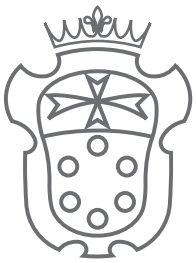


SCUOLA
NORMALE
SUPERIORE

Ph.D. in Chemistry

Computational methods for spectroscopic properties

Candidate
Ivan Carnimeo

Supervisor
Prof. Vincenzo Barone

Academic Year 2013-2014

II

Contents

Introduction

Vibrational spectroscopic experiments are nowadays largely employed in many applications, due to the richness of information that can be obtained about the structural properties and the chemico-physical processes occurring in the chemical systems. At the same time, the analysis and interpretation of such experimental data can be rather intricate, due to the large amount of effects of different typology and nature usually shuffled into the real systems. In fact, even in cases of small molecules in the gas-phase,[1–5] the interaction between the species in the sample can be rather complex, and can strongly affect the features of the spectra. Analogously, the spectra of flexible molecules can be composed by separate contributions originating from the different conformers, being of difficult interpretation.[4,5] An additional source of complexity also originates from the environmental effects, such as the interaction between the sample and the solvent[6,7] or the matrix medium.[5] As far as the dimension and the complexity of the molecular systems increase, the analysis of the experimental data become more difficult. In such circumstances, *in silico* modeling is a useful tool in order to assist the interpretation of the experimental measurements, due to the capability to predict the properties of the molecular systems on the basis of chemical-physical models. This can be interesting also in cases in which the sampling is not practical, as e.g. when the atmospherical or astrochemical processes are investigated,[8] or in case of the chemical analysis of pieces of cultural and artistic value, in which the sampling procedures produce permanent damages to the artworks.

The simulation of vibrational spectra requires the solution of the complete (electronic and nuclear) Schrödinger equation for polyatomic molecules, and the treatment of the light-matter interaction in an effective way, in order to determine the position and intensities of the peaks, respectively.[9–15] Although both time-dependent and time-independent routes have been proposed to this end,[11] in this work we will refer only to time-independent approaches in view of their efficiency and ease of use. Within the Born-Oppheimer approximation, this is accomplished by solving the electronic problem first, and then treating the nuclear motion in the force field generated by the electrons. Then, once the molecular states are known, the interaction between the molecular wavefunction and the incident field can be computed. As a consequence, in order to achieve a good (possibly quantitative) agreement between the simulated and the experimental spectra, both the nuclear (i) and the electronic

(ii) Schrödinger equations must be solved with an high accuracy.

Regarding the treatment of the nuclear problem (i), first the intramolecular potential must be defined, and then either variational or perturbative procedures can be applied. The intramolecular potential can be expressed using different sets of coordinates. The most common choice is to represent it as a function of the normal coordinates, obtained by diagonalizing the Hassian matrix expressed in mass-weighted cartesian coordinates.[9–11] Since both the normal and the cartesian coordinates depend on the absolute position of the molecule in the space, the six degrees of freedom associated to the rotational and vibrational motions of the molecule must be removed from treatment. Another possible choice is the use of the internal coordinates, which are able to describe the vibrations of the molecule being unaffected by the rotational and translational motions.[9,10] However, general and automated procedures able to define the internal coordinates without redundancies and for a wide range of molecules are still lacking, so that the definition of such coordinates requires a system-dependent analysis. For this reason the normal coordinates are usually preferred.

Using the normal coordinates, the intramolecular potential can be expressed in different functional forms. One approach worth of mention is the mode coupling expansion,[16,17] in which the potential is expressed as a series of functions of increasing dimensionality. At the first order, functions depending on the coordinates of only one mode are used, in order to take into accout the intrinsic anharmonicity of the normal modes; at the second order the coupling between pairs of modes is included by using functions depending on the coordinates of two normal modes; at the higher orders multi-dimensional functions are used. Usually such a form of the potential is truncated at the second order,[18–21] in order to reduce the computational demand required by the evaluation of the Potential Energy Surface (PES). Another very common representation for the intramolecular potential is based on a fourth-order Taylor expansion on the normal coordinates[12,22–29] and the numerical evaluation of the third and semi-diagonal fourth derivatives of the PES is required. Being the analytical second derivatives of the energy usually available for the most common electronic methods, the higher-order derivatives can be numerically evaluated in a robust way. Such a representation of the potential is suitable for small vibrations, while when the vibrational amplitude is large it can result in a poor representation of the PES.

Once the intramolecular potential is defined, the corresponding vibrational Schrödinger equation must be solved. A widely used approach is based on the vibrational self-consistent field (VSCF[16,17,19,21,30–35]) employing potentials in normal coordinates, based on either the mode coupling or Taylor expansions. Next, correlation can be taken into account by vibrational Møller-Plesset perturbation theory (VMP, also known as correlation-corrected VSCF, cc-VSCF),[36,37] vibrational configuration interaction (VCI),[38,39] or vibrational coupled clusters (VCC).[40] Many effective applications of the variational schemes have been reported[18–21] and some of them will also

be discussed in this work. While such procedures represent a promising route for the calculation of the anharmonic frequencies, the use of the vibrational second-order perturbation theory (VPT2) based on the Taylor expansion of the potential with respect to the normal modes[12,22–29] remains very attractive at least for semi-rigid systems.[6,29,41,42] Furthermore, the recent development and implementation of automatic and system-independent schemes[24–26,43] able to treat the singularities that usually affect the perturbative approaches, has sensibly increased the range of applicability of the perturbative treatment of the anharmonicity, including also those typology of molecules heavily plagued by the Fermi resonances. In particular, the HDCPT2[25] scheme is particularly suitable when large systems are involved.

Many methods have been proposed in order to simplify the treatment of the nuclear motion eventually reducing the corresponding computational burden. Many of them actually act on the harmonic part of the calculations by reducing the dimensionality of the Hessian matrix. The basic idea is to partition the molecular system in blocks, performing the geometry optimization and the Hessian calculations only on a part of the system.[44] The main issue connected with this procedure is that it leads to wrong and most likely imaginary frequencies, so that a certain number of approaches can be applied in order to perform the normal mode analysis on the partially optimized structure, ranging from simply assign an infinite mass to the frozen atoms,[45,46] to the inclusion of the coupling between the frozen and the mobile part in an approximate fashion (see Refs.[44,47–49] and references therein).

Approaches based on the neglection of a part of the derivatives can be also applied in the VPT2 treatment. In particular, we deloped an approximate VPT2 approach (RD-VPT2[50]), in which some of the third and fourth order derivatives of the PES were neglected, allowing to successfully simulate the vibrational spectrum of the glycine molecule on the (100) Silicon surface,[51] as will be discussed in detail in Chapter 3.

Regarding the electronic Schrödinger equation (ii) many methods are currently available for accurate force field evaluations. Approaches based on post Hartree-Fock (HF) theories are usually the methods of election when the accuracy of the results is a mandatory priority. For example, they have been extensively used for the computation of vibrational frequencies[26,42,52–55] and electronic excitation energies[56–60] and a quantitative agreement with the experimental measurements is nearly always obtained, provided that correlation and basis set completeness are taken into account in a proper and balanced way. Analogously, those methods have been used in composite approaches such as the "high accuracy estrapolated ab initio thermochemistry" (HEAT) protocol,[61–63] for computation of the atomization energies and other thermodynamical properties in quantitative agreement with the most accurate calorimetric measurements.[64] The main drawback of such methods is the large computational time and memory required by the calculations, being are also poorly scaling when the dimension of the systems increases. For this reason, approximate approaches must be employed for the solution of the electronic problem in the case

of large systems, and force fields based on the Møller-Plesset (MP2[65]) theory or single-determinant methods, such as the HF, Density Functional Theory (DFT) or semi-empirical methods, must be used. Among these possibilities, a particularly effective scheme is obtained when the PES at the DFT level is used, employing hybrid (especially B3LYP[66]) or double-hybrid (especially B2PLYP[67–69]) functional with medium-sized (double- or triple-ζ, polarization augmented) basis sets.[6,70–72] Hybrid force fields based on the use of two different methods, one – more accurate – for the evaluation of the harmonic part, the other – less computationally demanding – for the calculation of the higher order derivatives,[6,70,71,73–77] can be also used. Hybrid force fields will be used in Chapter 2 in order to accurately simulate the anharmonic spectra of small molecules, and in Chapter 5 for the inclusion of the anharmonic effects in the vibrational spectra of larger molecules.

Unfortunately, in many cases the spectroscopic processes primarly induced by a well localized chromophore are also significantly affected by the presence of more distant atoms (the "environment"), which should be thus included in the computational models. In this cases, the usual force fields based on the MP2 or DFT methods are not practical, being the computation of the required derivatives too demanding. Then, a next level of approximation in the solution of the electronic Schrödinger equation is the use multi-layered methods in order to treat only a reduced part of the system – where the chromophore is localized – at an higher level of theory, while taking into account the tuning by the remaining part of the system at a lower level. This is especially useful in the case of spectroscopic processes, whose simulation requires high accuracy and the involvement of the excited states, and is typically performed using quite costly and poorly scaling approaches.[11,74,78,79] Several multi-layered methods e.g. based on the ONIOM paradigm[80–86] and integrating quantum mechanics and molecular mechanics (QM/MM) models have been introduced. Such schemes proved remarkably successful and are also being further complemented by continuum models to deal with bulk solvent or matrix effects.[87–90]

Solvent effects can be accounted for in an implicit way, i.e. considering the solvent as a continuum interacting with the solute charge distribution in a pure electrostatic picture, or alternatively an explicit solvent model can be used, where e.g. a QM method is employed for the description of the solute in conjunction with a less accurate and less costly method applied to the solvent molecules. An intermediate approach in which the solute and few explicit solvent molecules are embedded in a continuum polarizable solvent model, has also been shown to achieve highly accurate results by including the polarization effects of the bulk of the solvent media in the most cost-effective way (see, for example Improta et al.[91]). Among the implicit solvent models, the Polarizable Continuum Model (PCM) in all its different variants (C-PCM,[92,93] D-PCM,[94] IEF-PCM[95–97]) is one of the most reliable methods, and it has been successfully applied for the simulation of many different properties of molecules in condensed phase. Moreover, non-equilibrium polarization effects[98–100] can be easily accounted for leading to a great improvement of the accuracy of the

predicted spectroscopic properties.[98–100]

A further degree of freedom is offered by the inclusion of an intermediate (QM') level between the QM and the MM ones, employing a semiempirical method which has the ability of describing genuine quantum effects at a reduced computational cost and with acceptable accuracy. Once again, this extension is especially significant when the dimension of the core region is limited by the cost of the most accurate method, but the tuning effect of the microscopic environment can go beyond simple electrostatic or polarization contributions. The first semi-empirical methods were proposed in order to reduce the computational time required by the evaluation of all the bielectronic integrals of the Hartree-Fock scheme. Such methods were based on the application of the zero differential overlap (ZDO) approximation to the Fock matrix written in atomic orbital basis set. On the basis on how many and which integrals were neglected, and on the parametrization schemes, different semiempirical methods have been developed, and among the most popular we found the CNDO,[101–103] INDO,[104] AM1[105] and PM6.[106] Originarly, such methods were used for the calculations of the electronic structures of molecules such as small alkanes,[107,108] but with the increasing of the performances of the computing machines, far from being abandoned, the semi-empirical methods started to be applied to the treatment of larger systems,[20,109,110] possibly within QM/QM', QM/QM'/MM or QM/QM'/PCM schemes. For this reason, the interest on the approximate electronic structure methods is still high and semi-empirical methods based on the DFT have also been developed.

The Self Consistent Charge Density Functional Tight Binding[111,112] (SCC-DFTB, or simply DFTB for conciseness) is a semi-empirical method particulary effective for the treatment of systems composed by a large number of atoms with a relative high accuracy. Directly derived[113–115] from the Density Functional Theory (DFT), it retains many characteristics of a fully QM method and in some cases it has been found to be a valuable alternative to the other semiempirical methods based on the wavefunction theory.[116–118] In particular, a very interesting point is the remarkable accuracy of the DFTB method, with respect to the other semi-empirical ones for the calculation of vibrational properties. Over a set of 66 organic molecules, Witek et al.[117] found an overall mean absolute deviation (MAD) of 57 cm^{-1} for SCC-DFTB harmonic frequencies, with respect to fundamental frequencies from experiment. Therefore, they proposed an improved parametrization of the repulsive potential for the C-C, H-H and C-H interactions, which reduced the MAD from 59.0 to 33.2 cm^{-1} over a set of 14 small hydrocarbon molecules.[119] Note however that this result was accomplished at the price of a reduced accuracy in equilibrium geometries and atomization energies. Automated parametrization procedures have been developed,[120,121] in order to fit the repulsive potentials to selected molecular data. A MAD of 33 and 38 cm^{-1} with respect to experimental fundamentals has been reported for hydrocarbon molecules[120] using two sets of parameters, developed with automated procedures. Although such approaches allowed to simulate IR[122]

and Raman[123] spectra of several molecules involving only C and H atoms (hydrocarbons, fullerenes,[124] nanodiamonds,[123] nanotubes[122]), it cannot be easily translated into a general purpose protocol to improve the accuracy of predicted spectroscopic properties, mainly because the contributions from the higher derivatives of the potential are completely neglected in the computation of vibrational frequencies. A different yet complementary approach to vibrational spectroscopies follows a time-dependent route, employing Fourier transforms of the autocorrelation functions of dipole moments[125,126] and polarizabilities[125] accumulated during molecular dynamics simulations. In the implementation proposed by Elstner et al.,[125] large deviations were found in the position of the peaks in the Raman spectrum, such as the ones related to the O-H stretching modes of water (error of 244 cm^{-1} with respect to B3LYP/AVTZ calculations), to the C-H stretching region of glycerol (error of $\sim$ 160 cm^{-1} with respect to BLYP/AVTZ), or to the S-H stretching mode of ethanethiol (error of $\sim$ 130 cm^{-1} with respect to BLYP/AVTZ). Large errors have also been found in the position of the C-C and C-H stretchings, which have been corrected using the set of parameters developed by Witek et al.[119,120] A time-dependent approach has also been used by Spiegelman et al.,[126] for the study of water-benzene complexes. In this case, the standard DFTB method has been improved with the inclusion of an empirical correction[127,128] to account for dispersion interactions, and with the use of CM3 charges[128–130] in place of Mulliken charges, to improve the treatment of the bond polarization effects on the IR intensities.

Other developments of the DFTB methods have been done in our group. First of all, an analytic version of the method (DFTBA[131]) has been proposed, in order to compute the molecular properties, thanks to a robust implementation of the analytical derivatives of the interaction parameters. Then, the integration with the PCM in an effective DFTB/PCM scheme has been implemented[132] in the GAUSSIAN code,[133] allowing to include the solvent effects in the DFTB computations in an implicit way. Finally, the VPT2 method complemented with the schemes for the treatment of the Fermi resonances outlined above, has been used together with an hybrid DFT/DFTB/PCM force field for the calculation of the anharmonic frequencies of Uracil in aqueous solution. The latter system has been chosen as a test case, and the good accuracy of the results obtained is encouraging toward the effective implementation of multi-layered QM/QM'/PCM, or QM/QM'/MM/PCM approaches for the treatment of large complex systems in solution. It is worth noting that although the present work will mainly deal with the computation of vibrational properties, developments on the calculation of the vertical excitation energies has been also performed, with the implementation of the TD-DFTB[134] and TD-DFTB/PCM[132] schemes.

In this work the main results obtained during my Ph.D. and published in various articles[50,51,70,71,73,132,134] in the past years have been summarized, the guideline connecting the different works having been evidenced. The reader interested in more

details about the topics discussed in this thesis is referred to the original articles, which have been cited in the relevant points and have been also listed in chronological order in the List of Articles at the bottom of this thesis. In Chapter 1 the main aspects of the theoretical background of the VPT2 approach, together with the most effective schemes for the treatment of Fermi resonances are briefly illustrated. Then, the discussion of two specific applications of the method to the study of molecules of atmospherical interest[70,73] and to the conformational analysis of a gas-phase glycine mixture,[71] and of one benckmark work validating the accuracy of the VPT2 frequencies and intensities[70] will follow. In Chapter 2 the basic theory and results[70,71,73] of the computation of anharmonic frequencies and intensities using hybrid methods are illustrated. Chapter 3 is focused on the reduced dimensionality VPT2 approach. At the beginning of the Chapter the theoretical background is presented,[50] while at the end the simulation of the HREEL spectrum of the glycine molecule adsorbed on the (100) Silicon surface is discussed.[51] In the last two Chapters an accurate description of the DFTB/PCM[132] and TD-DFTB/PCM[132,134] schemes is reported. In Chapter 4 the theory and implementation in the GAUSSIAN code[133] of the DFTB, DFTB/PCM and TD-DFTB/PCM approaches are presented and discussed. Then, in Chapter 5 along with the benchmark tests of the DFTB/PCM shemes, different applications are proposed, regarding the accuracy of the TD-DFTB/PCM scheme for the computation of vertical excitation energies of Uracil and Red Nile in solution, and the calculation of the anharmonic vibrational frequencies of Uracil in water, using the hybrid approaches described in Chapter 2.

Chapter 1

Theoretical modeling of vibrational spectra

Infrared spectra are generally measured in terms of *absorbance*, which can be related to the molar absorption coefficient following the Beer-Lambert law

$$\mathcal{A} = \varepsilon(\nu) l C \tag{1.1}$$

where C is the concentration of the sample, usually expressed in [mol L^{-1}], and l is the the optical pathlength in cm. The spectral features are determined by an intrinsic molecular property, function of the frequency of the incident light (ν), usually referred to as the *molar absorption coefficient* (ε) when expressed in macroscopic quantities, or as the *absorption cross section* (σ), microscopic counterpart[a]

$$\sigma \left[\text{cm}^2 \text{ molecule}^{-1}\right] = \frac{N_A}{10^3 \ln(10)} \cdot \varepsilon \left[\text{L mol}^{-1} \text{ cm}^{-1}\right] \tag{1.2}$$

with N_A the Avogadro constant (mol^{-1}). The *absorption cross section* is the result of the physico-chemical processes induced by the light-matter interaction, which in the case of the vibrational spectra lead to many complex low-intensity features, such as overtones and combination bands, resonances, or to rotational substructure of the vibrational bands. Consequently, in order to provide a reliable quantification of the intensity of the transitions, the experimental value of $\varepsilon(\nu)$ is obtained through a numerical integration over a chosen range of frequencies, e.g. $[\nu_1, \nu_2]$,

$$G_{\nu_1,\nu_2} = \int_{\nu_1}^{\nu_2} \varepsilon(\nu) d\nu. \tag{1.3}$$

The *integrated absorption cross section*[1–3] ($G_{\nu_1\nu_2}$), is thus obtained as the area spanning over the region of the spectrum in the frequency interval $[\nu_1, \nu_2]$, and is usually given in km/mol. The quantum mechanical description of the light-matter

[a]The natural logarithm will be used here, although the decadic logarithm is sometimes used for the absorbance when dealing with solvents.

interaction provides a physical interpretation of the phenomena contributing to the observed *absorption cross section*, eventually allowing to compute a theoretical estimation which can be directly compared to the experimental one. The underlying theory can be rather complex and will only be briefly presented here, focusing on the main concepts. In the following, the main steps of the derivation will be presented, starting from the definition of the Hamiltonian to the description of effective computational schemes able to quantitatively predict the *absorption cross sections* of the Infrared spectra. A much more complete derivation with further details can be found elsewhere.[9–15] At the end of the Chapter, some numerical tests are presented, and one application is discussed.

1.1 The vibrational perturbative approach

1.1.1 The interaction Hamiltonian

It can be shown[10] that the Hamiltonian of a molecular system embedded in a radiation field of M photons, can be written as

$$\mathcal{H} = \mathcal{H}_{mol} + \mathcal{H}_{field} + \mathcal{H}_{int} \tag{1.4}$$

where $\mathcal{H}_{mol}$ and $\mathcal{H}_{field}$ are the Hamiltonians of the unperturbed molecule and radiation field, respectively, while $\mathcal{H}_{int}$ describes the interaction between the molecule and the field. The interaction Hamiltonian can be written in terms of the photon-creation (a^*_λ) and photon-annihilation (a_λ) operators

$$\begin{aligned}\mathcal{H}_{int} = &\sum_{\lambda}^{M} H^{(1)}_\lambda(a_\lambda) + H^{(1)}_\lambda(a^*_\lambda) + \\ &+ \sum_{\lambda\mu}^{M} H^{(2)}_{\lambda\mu}(a_\lambda, a_\mu) + H^{(2)}_{\lambda\mu}(a^*_\lambda, a_\mu) + H^{(2)}_{\lambda\mu}(a_\lambda, a^*_\mu) + H^{(2)}_{\lambda\mu}(a^*_\lambda, a^*_\mu)\end{aligned} \tag{1.5}$$

and it is general enough to describe all the one- and two-photon processes induced by the light-matter interaction. The one-photon operators $H^{(1)}_\lambda(a_\lambda)$ and $H^{(1)}_\lambda(a^*_\lambda)$, and the two-photons operators $H^{(2)}_{\lambda\mu}(a_\lambda, a_\mu)$ and $H^{(2)}_{\lambda\mu}(a^*_\lambda, a^*_\mu)$, describe the absorption and emission spectroscopies, while $H^{(2)}_{\lambda\mu}(a^*_\lambda, a_\mu)$ and $H^{(2)}_{\lambda\mu}(a_\lambda, a^*_\mu)$ are related to the Raman spectroscopy. In this work, we are mainly interested in the one-photon Infrared absorption spectroscopy, which is connected to the $H^{(1)}_\lambda(a_\lambda)$ operator, so that $\mathcal{H}_{int}$ takes the form

$$\mathcal{H}_{int} = \sum_{\lambda}^{M} H^{(1)}_\lambda(a_\lambda) = -\frac{2\hbar\pi^{\frac{1}{2}}}{i} \sum_{j}^{N_p} \sum_{\lambda}^{M} \frac{e_j}{m_j} a_\lambda \exp\left[\frac{i}{\hbar c}(\mathbf{k}_\lambda \cdot \mathbf{r}_j)\right] (\mathbf{u}_\lambda \cdot \nabla_j). \tag{1.6}$$

The operator in Eq. 1.6 describes the annihilation of one photon of energy $\hbar\omega_\lambda$ due to the interaction with a molecular system composed by N_p particles of charges e_j

and masses m_j (including both the nuclei and the electrons). $\mathbf{k}_\lambda$ is the direction of propagation of the photon λ, $\mathbf{r}_j$ is the position vector of the particle j, and $\mathbf{u}_\lambda$ is a unit vector perpendicular to the direction of propagation of the photon, so that the scalar product $\mathbf{u}_\lambda \cdot \nabla_j$ takes into account the relative orientation between the photon and the position of the particle.

The time-evolution of such a system under irradiation is determined by the time-dependent Schrödinger equation

$$\mathcal{H}\Phi(t) = i\hbar\frac{\partial\Phi(t)}{\partial t}, \tag{1.7}$$

where $\Phi(t)$ is the time dependent wavefunction of the whole system, describing both the molecule and the field states. Assuming that the eigenfunctions of the unperturbed molecule and field are known, Eq. 1.7 can be solved with the time-dependent perturbation theory.[10] As a result, the probability that a transition occurs, from an initial molecular state $|\Psi_I\rangle$ to a final molecular state $|\Psi_F\rangle$, as a consequence of the absorption of one photon, is obtained as

$$w_{IF} = \frac{2\pi}{3\hbar}|\langle\boldsymbol{\mu}\rangle_{IF}|^2\rho(E), \tag{1.8}$$

where $\rho(E)$ is the energy density of the field, a quantity connected to the intensity of the beam of the incident light. The key quantity $\langle\boldsymbol{\mu}\rangle_{IF}$ is the electric transition dipole moment, defined as

$$\langle\boldsymbol{\mu}\rangle_{IF} = \langle\Psi_I|\boldsymbol{\mu}|\Psi_F\rangle \tag{1.9}$$

with $\boldsymbol{\mu}$ the electric dipole moment operator

$$\boldsymbol{\mu} = \sum_j^{N_p} e_j\mathbf{r}_j. \tag{1.10}$$

In the following section we will present the main steps leading to the solution of the molecular Schrödinger equation connected to $\mathcal{H}_{mol}$, in order to find the eigenfunctions $|\Psi_I\rangle$ and $|\Psi_F\rangle$ of the unperturbed molecule, required for the evaluation of the integrals of the type of Eq. 1.9.

1.1.2 The molecular Hamiltonian

Given a molecular system composed of nuclei and electrons, the non-relativistic molecular Hamiltonian $\mathcal{H}_{mol}$ can be written in terms of the nuclear ($\mathbf{R}$) and electronic ($\mathbf{r}$) coordinates, the molecular wavefunction being $\Psi(\mathbf{R},\mathbf{r})$. Within the Born-Oppenheimer approximation,[10,135] the molecular wavefunction is factorized as $\Psi(\mathbf{R},\mathbf{r}) = \Psi_e(\mathbf{R},\mathbf{r})\Psi_n(\mathbf{R})$, and the electronic problem is solved by considering the electrons moving in a field of fixed nuclei, so that the electronic energy $\mathcal{V}$ is a function of $\mathbf{R}$. Then, the nuclear Hamiltonian is

$$\mathcal{H}_n = \mathcal{T}_n + \mathcal{V}, \tag{1.11}$$

where $\mathcal{T}_n$ is the kinetic energy operator associated to the nuclear motion and the electronic energy is the potential field, usually called the potential energy surface (PES) of the molecule. In order to effectively solve the nuclear problem, an explicit form for the nuclear coordinates $\mathbf{R}$ should be provided, by defining a reference frame of axes. Such a frame is chosen in order to describe the translational, vibrational and rotational motions of the molecule, and the *Sayvez conditions* are employed to decouple the three contributions in the kinetic energy. The nuclear Hamiltonian can be then written in the form[9,10]

$$\mathcal{H}_n = \mathcal{T}_t + \mathcal{T}_r + \mathcal{T}_{Coriolis} + \mathcal{T}_v + \mathcal{V} \tag{1.12}$$

where $\mathcal{T}_t$, $\mathcal{T}_r$ and $\mathcal{T}_v$ are the translational, rotational and vibrational kinetic energies, respectively, and $\mathcal{T}_{Coriolis}$ is the vibro-rotational coupling. The latter term couples the vibrational and rotational motions of the molecule, and is not due to the particular choice of the coordinates but rather to a physical effect. Since it is a small contribution to the energy it can be neglected for most purposes. Detailed treatments of the vibro-rotational Hamiltonian in both the classical and quantum mechanics have been presented in many books by several authors.[9–15,136,137]

Since we are mostly interested in the vibrational spectroscopy, we will focus on the vibrational Hamiltonian, $\mathcal{H}_v = \mathcal{T}_v + \mathcal{V}$, in the following. Usually, the mass-weighted normal coordinates $\mathbf{Q}$ are used to describe the vibrational motion, although this is not a fixed rule and other sets of coordinates (such as the internal) can be used. A remarkable point is that for a non-linear[b] molecule composed of N atoms and uniquely identified by $3N$ cartesian coordinates, only $3N-6$ coordinates (either normal or internal) are required for the complete decription of the vibrational motion, since 6 degrees of freedom are associated to the traslational and rotational motions. Using the normal coordinates, for a non-linear molecule composed of N atoms $\mathcal{H}_v$ can be written as

$$\mathcal{H}_v(\mathbf{Q}) = -\frac{1}{2}\sum_{i}^{3N-6}\frac{\partial^2}{\partial Q_i^2} + \mathcal{V}(\mathbf{Q}). \tag{1.13}$$

In order to solve the eigenvalue equation for such an Hamiltonian a functional form for $\mathcal{V}(\mathbf{Q})$ should be provided. The simplest approach is to use an harmonic potential, so that Eq. 1.13 is reduced to the equation of the harmonic oscillator, which can be solved analytically. Then, the set of eigenstates can be easily obtained, the vibrational wavefunction, $\Psi_v(\mathbf{Q})$, being a product of $3N-6$ independent oscillator wavefunctions, $\varphi_i(Q_i)$

$$\Psi_v(\mathbf{Q}) = \prod_{i}^{3N-6}\varphi_i(Q_i) \tag{1.14}$$

[b]For linear molecules only 2 degrees of freedom are associated to the rotational motion, so that the vibration is described by $3N-5$ coordinates.

written in terms of the *Hermite polynomials*. The expectation value of Eq. 1.13 then reads

$$\langle \Psi_v | \mathcal{H}_v | \Psi_v \rangle = \sum_i^{3N-6} \omega_i \left(n_i + \frac{1}{2} \right), \tag{1.15}$$

where ω_i and n_i are respectively the harmonic frequency and the vibrational quantum number associated to the ith mode of vibration. Once the solutions of the harmonic oscillator are known, the harmonic frequencies can be used to estimate the peak positions, while the Infrared spectrum can be computed by evaluating the integrals of type of Eq. 1.9. In particular, the dipole moment operator is expanded in Taylor series with respect to the normal coordinates truncated at the first order (*electric* harmonicity), and the harmonic wavefunctions are used to describe the initial and the final states of the vibrational transition (*mechanical* harmonicity). Unfortunately, the harmonic energies usually provide a poor estimation of the experimental peak positions, while when the *double harmonic* approximation is employed, non-zero intensities are obtained only for the fundamental transitions, due to the orthogonality of the *Hermite polynomials*. For these reasons, in order to accurately compute Infrared spectra it is important to go beyond the harmonic approximation for both the frequencies and the intensities (see also Section 1.1.5).

1.1.3 Anharmonicity

On the basis of the functional form chosen for the intramolecular potential and on the subsequent treatment of the vibrational Schrödinger equation, several approaches can be distinguished in order to go beyond the harmonic approximation. Regarding the choice of $\mathcal{V}$, among the different possibilities, the polynomial representation is one of the most used forms. In particular, the potential is expanded in Taylor series on the normal coordinates about the equilibrium geometry[c],[10,11,24]

$$\mathcal{V} = \frac{1}{2} \sum_i^{3N-6} \omega_i q_i^2 + \frac{1}{6} \sum_{ijk}^{3N-6} k_{ijk} q_i q_j q_k + \frac{1}{24} \sum_{ijkl}^{3N-6} k_{ijkl} q_i q_j q_k q_l + ... \tag{1.16}$$

In Eq. 1.16, the first term is the harmonic potential, while k_{ijk} and k_{ijkl} are the third and fourth derivatives of the PES over the "dimensionless" normal coordinates $\mathbf{q}$ at the equilibrium geometry, also referred to as the cubic and quartic force constants

$$k_{ijk} = \left(\frac{\partial^3 \mathcal{V}}{\partial q_i \partial q_j \partial q_k} \right)_0 \quad \text{and} \quad k_{ijkl} = \left(\frac{\partial^4 \mathcal{V}}{\partial q_i \partial q_j \partial q_k \partial q_l} \right)_0 . \tag{1.17}$$

[c]The first order term in Eq. 1.16 is zero since the first derivatives computed at the equilibrium geometry are zero by definition, while the zeroth order term is set to zero ($\mathcal{V}_0 = 0$) by choosing the electronic energy at the equilibrium geometry as the reference for the energy.

The dimensionless coordinates $\mathbf{q}$ can be related to the mass-weighted normal coordinates $\mathbf{Q}$ by the relation

$$q_i = \left(\frac{\hbar}{2\pi c\omega_i}\right)^{\frac{1}{2}} Q_i \tag{1.18}$$

and such a set of coordinates is introduced in order to allow direct comparison between the force constants at different orders, which otherwise would have different dimension.[d] Although such a representation of the potential is not suited for the treatment of the modes with large amplitude, being affected by truncation errors, it has the strong advantage to be very simple and general enough to be applied to a wide range of systems, retaining an high accuracy for most of the purposes.

Whenever the terms higher than the quadratic one are retained in Eq. 1.16, the eigenvalue equation associated with the operator in Eq. 1.13 can not be solved analytically, so that variational or perturbative methods are usually applied. A widely used approach is based on the variational self consistent field (VSCF[16,17,19,21,30–35]), and solves the vibrational Schrödinger equation in a similar way as the Hartree-Fock method[135] does for the electronic problem. The ansatz for the vibrational wavefunction is based on the harmonic oscillator eigenstates, approximating the potential energy with a sum of mean field potentials, acting on each normal mode. Once the VSCF has been performed, the correlation between the normal modes can be taken into account with post-SCF treatments, such as the vibrational Møller-Plesset perturbation theory (VMP, also known as correlation-corrected VSCF, cc-VSCF),[36,37] the vibrational configuration interaction (VCI),[38,39] or the vibrational coupled-cluster (VCC).[40]

Another approach is based on the partition of the intramolecular potential, as represented in Eq. 1.16, in different orders of perturbation,

$$\mathcal{V} = V^{(0)} + \lambda V^{(1)} + \lambda^2 V^{(2)} + ... \tag{1.19}$$

where λ is an arbitrary perturbation parameter. Then, the vibrational Schrödinger equation is solved using the standard Rayleigh-Schrödinger Perturbation Theory[135] (RSPT), treating the harmonic Hamiltonian as the zeroth-order operator, and the terms depending on the third- and fourth-order derivatives of the PES as first and second order perturbations.[e] Such an approach is usually referred to as the VPT2 approach.[10,11,23–25,141–143] The VPT2 approach has been extensively used in this work, due to its cost-effectiveness and flexibility for the extension to large systems and to the reliability of the results which has been reported in many cases.[6,51,70–75,144] In the version proposed by Barone[23,24] and implemented in the GAUSSIAN code,[133]

[d] The dimensionless normal coordinates are also useful when the second quantization formalism is used.

[e] Although the relevant equations were first formulated by Miller *et al.*[138] in terms of the Van Vleck[139] perturbation theory, the derivation based on the RSPT is often used due to its compactness and relative simplicity of the equations.[140]

the VPT2 treatment is applied to the complete vibro-rotational Watson Hamiltonian,[136,137] so that a fraction of the Coriolis coupling operator ($\mathcal{T}_{Coriolis}$) is included in $V^{(2)}$

$$\sum_{\tau=a,b,c} B_\tau^e \left[\sum_{i<j} \zeta_{ij}^\tau (q_i p_j - q_j p_i)\right]^2, \tag{1.20}$$

where p_i is the conjugate momentum of q_i, τ identifies the rotational axes of the molecule and ζ_{ij}^τ and B_τ^e are the Coriolis coupling and the equilibrium rotational constants, respectively.

Then, for a vibrational state defined by

$$|\Psi_v\rangle = |n_1, n_2, ..., n_{3N-6}\rangle \tag{1.21}$$

where n_i is the quantum number associated to the i-th mode of vibration, the VPT2 energy can be expressed as the expectation value of an effective Hamiltonian[24] ($\mathcal{H}^{\text{eff}}$)

$$\langle\Psi_v|\mathcal{H}^{\text{eff}}|\Psi_v\rangle = \chi_0 + \sum_i \omega_i \left(n_i + \frac{1}{2}\right) + \sum_{i\leq j} \chi_{ij} \left(n_i + \frac{1}{2}\right)\left(n_j + \frac{1}{2}\right) \tag{1.22}$$

The higher order derivatives of the PES – i.e., the anharmonic contributions – are included in χ_0 and in the χ_{ij} matrix elements,[24,138] which for an asymmetric top are defined as

$$\begin{aligned} 64\chi_0 = &\sum_i k_{iiii} - \frac{7}{9}\sum_i \frac{{k_{iii}}^2}{\omega_i} + 3\sum_{\substack{ij\\ j\neq i}} \frac{\omega_i k_{ijj}^2}{(4\omega_j^2 - \omega_i^2)} \\ &- 16\sum_{\substack{ijk\\ k>j>i}} \frac{\omega_i\omega_j\omega_k k_{ijk}^2}{\Delta_{ijk}} \quad - 16\sum_{\tau=a,b,c} B_\tau^e \left[1 + 2\sum_{\substack{ij\\ j>i}} (\zeta_{ij}{}^\tau)^2\right] \end{aligned} \tag{1.23}$$

$$16\chi_{ii} = k_{iiii} - \frac{5{k_{iii}}^2}{3\omega_i} - \sum_{\substack{j=1\\ j\neq i}} \frac{\left(8{\omega_i}^2 - 3{\omega_j}^2\right){k_{iij}}^2}{\omega_j\left(4{\omega_i}^2 - {\omega_j}^2\right)} \tag{1.24}$$

$$\begin{aligned} 4\chi_{ij} = &k_{iijj} - \frac{2\omega_i {k_{iij}}^2}{\left(4{\omega_i}^2 - {\omega_j}^2\right)} - \frac{2\omega_j {k_{ijj}}^2}{\left(4{\omega_j}^2 - {\omega_i}^2\right)} \\ &- \frac{k_{iii}k_{ijj}}{\omega_i} - \frac{k_{jjj}k_{iij}}{\omega_j} \\ &+ \sum_{\substack{k=1\\ k\neq i,j}} \left[\frac{2\omega_k\left({\omega_i}^2 + {\omega_j}^2 - {\omega_k}^2\right){k_{ijk}}^2}{\Delta_{ijk}} - \frac{k_{iik}k_{jjk}}{\omega_k}\right] \\ &+ \frac{4\left({\omega_i}^2 + {\omega_j}^2\right)}{\omega_i\omega_j} \sum_{\tau=a,b,c} B_\tau^e \left(\zeta_{ij}^\tau\right)^2 \end{aligned} \tag{1.25}$$

$$\Delta_{ijk} = {\omega_i}^4 + {\omega_j}^4 + {\omega_k}^4 - 2\left({\omega_i}^2{\omega_j}^2 + {\omega_i}^2{\omega_k}^2 + {\omega_j}^2{\omega_k}^2\right)$$

Please note that the Δ_{ijk} employed here is consistent with Refs.,[25,29] and it is defined with opposite sign with respect to Martin *et al.*[26] In Eq. 1.23, 1.24 and 1.25, the terms in which differences between the harmonic frequencies occur at the denominator are plagued by singularities, and will be discussed in detail in Section 1.1.4.

From Eq. 1.22 energies of all vibrational states of the molecule can be calculated. The expectation value of the ground-state energy, i.e., the Zero Point Vibrational Energy (ZPVE), is associated to the vibrational state

$$|\Psi_v\rangle = |0_1, 0_2, ..., 0_{3N-6}\rangle \tag{1.26}$$

and thus reads

$$\langle 0_1, 0_2, ...|\mathcal{H}^{\text{eff}}|0_1, 0_2, ...\rangle = \chi_0 + \frac{1}{2}\sum_i \omega_i + \frac{1}{4}\sum_{i\leq j}\chi_{ij}. \tag{1.27}$$

The energy of one fundamental transition associated to mode k can be calculated as the difference between the energy of the state

$$|\Psi_v\rangle = |0_1, 0_2, ..., 0_{k-1}, 1_k, 0_{k+1}, ..., 0_{3N-6}\rangle \tag{1.28}$$

and the ZPVE, resulting in

$$\nu_{0,1_k} = \omega_k + 2\chi_{kk} + \frac{1}{2}\sum_{i\neq k}\chi_{ik} \tag{1.29}$$

Similarly, the vibrational energies of all overtones, combination bands and multi-quanta transitions can be calculated. The expression for the frequency of two- and three- quanta transitions are reported in the following

$$\begin{aligned}
\nu_{0,2_k} &= 2\omega_k + 6\chi_{kk} + \sum_{i\neq k}\chi_{ik} \\
\nu_{0,1_k1_l} &= \omega_k + \omega_l + 2\chi_{kk} + 2\chi_{ll} + 2\chi_{kl} + \frac{1}{2}\sum_{i\neq k,l}(\chi_{ik} + \chi_{il}) \\
\nu_{0,3_k} &= 3\omega_k + 12\chi_{kk} + \frac{3}{2}\sum_{i\neq k}\chi_{ik} \\
\nu_{0,2_k1_l} &= 2\omega_k + \omega_l + 6\chi_{kk} + 2\chi_{ll} + 3.5\chi_{kl} + \sum_{i\neq k}\chi_{ik} + \frac{1}{2}\sum_{i\neq l}\chi_{il} \\
\nu_{0,1_k1_l1_m} &= \omega_k + \omega_l + \omega_m + 2\chi_{kk} + 2\chi_{ll} + 2\chi_{mm} + 2\chi_{kl} + 2\chi_{km} \\
&\quad + \sum_{i\neq k}\frac{1}{2}\chi_{ik} + \sum_{i\neq l}\frac{1}{2}\chi_{il} + \sum_{i\neq m}\frac{1}{2}\chi_{im}
\end{aligned} \tag{1.30}$$

The position of the Infrared peaks are given by the energies of all the vibrational states of the molecule. Since they are independent from a particular spectroscopy,

they can be used for the analysis of many different techniques (e.g., the IR, 2D-IR, Raman, Vibrational Circular Dichroism or Electron Energy Loss Spectroscopy). For this reason, the computation of accurate vibrational frequencies is by far one of the most important challenges of computational spectroscopy.

The VPT2 approach can be applied also for the calculation of anharmonic intensities[28,138,140,145] The electric dipole moment operator is expanded in Taylor series with respect to the dimensionless normal coordinates, about the equilibrium geometry

$$\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \sum_i \boldsymbol{\mu}_i q_i + \frac{1}{2}\sum_{ij} \boldsymbol{\mu}_{ij} q_i q_j + \frac{1}{6}\sum_{ijk} \boldsymbol{\mu}_{ijk} q_i q_j q_k + ... \tag{1.31}$$

with

$$\boldsymbol{\mu}_i = \left(\frac{\partial \boldsymbol{\mu}}{\partial q_i}\right)_0 \text{ and } \boldsymbol{\mu}_{ij} = \left(\frac{\partial^2 \boldsymbol{\mu}}{\partial q_i q_j}\right)_0 \text{ and } \boldsymbol{\mu}_{ijk} = \left(\frac{\partial^3 \boldsymbol{\mu}}{\partial q_i q_j q_k}\right)_0 \tag{1.32}$$

Eq. 1.9 then becomes

$$\begin{aligned}\langle \boldsymbol{\mu} \rangle_{IF} =& \boldsymbol{\mu}_0 \langle \Psi_v^I | \Psi_v^F \rangle + \sum_i \boldsymbol{\mu}_i \langle \Psi_v^I | q_i | \Psi_v^F \rangle + \frac{1}{2}\sum_{ij} \boldsymbol{\mu}_{ij} \langle \Psi_v^I | q_i q_j | \Psi_v^F \rangle \\ &+ \frac{1}{6}\sum_{ijk} \boldsymbol{\mu}_{ijk} \langle \Psi_v^I | q_i q_j q_k | \Psi_v^F \rangle + ...\end{aligned} \tag{1.33}$$

Eq. 1.33 is usually truncated at the cubic term and the anharmonic wavefunctions are used for the vibrational states $|\Psi_v\rangle$. Then, after the numerical evaluation of $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\mu}_{ijk}$ the RSPT is applied[f], treating the terms depending on the first derivatives of the dipole moment ($\boldsymbol{\mu}_i$) as the zeroth-order operator, the terms depending on $\boldsymbol{\mu}_{ij}$ as the first order perturbation, and terms depending on $\boldsymbol{\mu}_{ijk}$ as the second order perturbation. Finally, anharmonic $\langle \boldsymbol{\mu} \rangle_{IF}$ are obtained as functions of ω_i, k_{ijk}, k_{ijkl}, $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\mu}_{ijk}$, for every $I \rightarrow F$ transition, and the transition probabilities w_{IF} can be evaluated. Assuming that the initial state of the transition is the ground state, and the final state is the first excited state of the vibrational mode i,

$$\begin{aligned}|\Psi_v^I\rangle &= |0_1, 0_2, ..., 0_{3N-6}\rangle \\ |\Psi_v^F\rangle &= |0_1, 0_2, ..., 0_{i-1}, 1_i, 0_{i+1}, ..., 0_{3N-6}\rangle\end{aligned} \tag{1.34}$$

[f]Similarly to the energies, the Van Vleck[139] perturbation theory, can be used also for the intensities as an alternative to the RSPT.

the explicit form of the transition dipole moment can be given

$$
\begin{aligned}
\langle\, \mu^a \,\rangle_{0,1_i} =& \frac{\mu_i^a}{\sqrt{2}} + \frac{1}{4\sqrt{2}} \sum_j \mu_{jji}^a - \frac{1}{8\sqrt{2}} \sum_{jk} k_{ijkk} \mu_j^a \left[\frac{1}{\omega_i + \omega_j} - \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right] \\
& - \frac{1}{8\sqrt{2}} \sum_{jk} \left\{ k_{ijk} \mu_{jk}^a \left(\frac{1}{\omega_i + \omega_j + \omega_k} - \frac{1}{\omega_i - \omega_j - \omega_k} \right) + \frac{2k_{jkk}}{\omega_j} \mu_{ji}^a \right\} \\
& + \frac{1}{2\sqrt{2}} \sum_{jk} \left(\sum_{\tau=a,b,c} B_\tau^e \zeta_{ik}^\tau \zeta_{jk}^\tau \right) \mu_j^a \left\{ \frac{\sqrt{\omega_i \omega_j}}{\omega_k} \left(\frac{1}{\omega_i + \omega_j} + \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right) \right. \\
& \left. - \frac{\omega_k}{\sqrt{\omega_i \omega_j}} \left(\frac{1}{\omega_i + \omega_j} - \frac{1-\delta_{ij}}{\omega_i - \omega_j} \right) \right\} \\
& + \frac{1}{16\sqrt{2}} \sum_{jkl} k_{ikl} k_{jkl} \mu_j^a \left\{ \frac{4\omega_j (\omega_k + \omega_l)(1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})}{({\omega_j}^2 - {\omega_i}^2)\left[(\omega_k + \omega_l)^2 - {\omega_i}^2\right]} \right. \\
& + \frac{(\omega_k + \omega_l)\left[(\omega_k + \omega_l)^2 - 3{\omega_i}^2\right] \delta_{ij}(1+\delta_{ik})(1-\delta_{il})}{\omega_i \left[(\omega_k + \omega_l)^2 - {\omega_i}^2\right]^2} \\
& \left. + \frac{4\omega_j (3\omega_k + 4\omega_i)(1-\delta_{ij})(1-\delta_{ik})\delta_{il}}{\omega_k ({\omega_j}^2 - {\omega_i}^2)(\omega_k + 2\omega_i)} \right\} \\
& + k_{ijk} k_{llk} \mu_j^a \left\{ \frac{\delta_{ij}}{\omega_i \omega_k} \left(1 + \frac{2\delta_{ik}\delta_{il}}{9} \right) + \frac{4\omega_j (1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})}{\omega_k \left({\omega_j}^2 - {\omega_i}^2\right)} \right. \\
& \left. + \frac{4\omega_j \delta_{ik}(1-\delta_{ij})}{\omega_i ({\omega_j}^2 - {\omega_i}^2)} \left(1 + \frac{2\delta_{ij}}{3} \right) \right\}
\end{aligned}
\tag{1.35}
$$

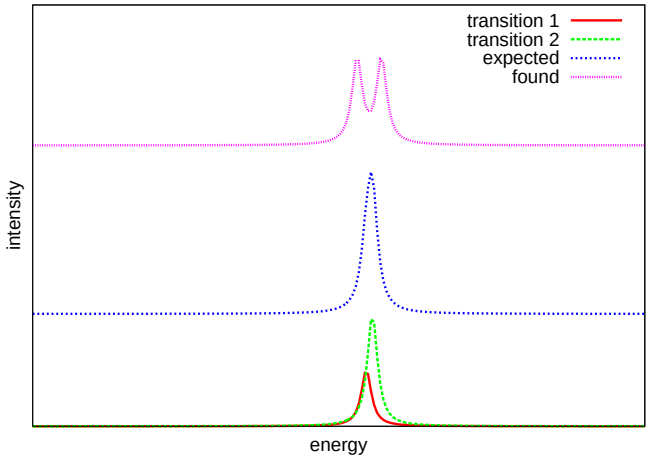
where μ^a is a cartesian component of the electric dipole and δ_{ij} is the Kronecker delta. The corresponding expressions for the non-fundamental transitions read

$$
\begin{aligned}
\langle\, \mu^a \,\rangle_{0,2_i} &= \frac{1}{2\sqrt{2}} \mu_{ij}^a + \frac{1}{4\sqrt{2}} \sum_k k_{ijk} \mu_k^a \left(\frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i + \omega_j + \omega_k} \right) \\
\langle\, \mu^a \,\rangle_{0,1_i 1_j} &= \frac{1}{2} \mu_{ij}^a + \frac{1}{4} \sum_k k_{ijk} \mu_k^a \left(\frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i + \omega_j + \omega_k} \right)
\end{aligned}
\tag{1.36}
$$

1.1.4 The treatment of the Fermi resonances

The phenomenon of the Fermi resonances is a physical effect strictly peculiar of the vibrational spectra, being directly connected to the purely vibrational Hamiltonian. In principle, when two transitions have similar energies they should appear very close in the spectrum resulting in one peak intense as the sum of the intensities of the two separate transitions. Contrarly, when the two transitions have proper symmetry properties it can be rather observed that they are well separated and have similar intensities.[10,146] Such an effect was recognized for the first time by Fermi in the

Raman spectrum of the CO_2 molecule,[147] and when it occurs it is usually said that the two transitions are resonant.



From a quantum mechanical perspective, the Fermi resonances can be considered as a mixing occurring between the final states of two transitions of similar energies, and they can be exactly described with a variational treatment. We imagine an Hamiltonian, $\mathcal{H}^*_{\text{eff}}$, which provides as expectation values the energies (in wavenumbers) of the two transitions, as if they were unaffected by the resonance. So they will be refferred to as the "deperturbed" frequencies hereafter and labelled as ν_1^* and ν_2^*. Then, assuming that a resonance occurs, the vibrational wavefunctions of the final states of the two transitions, $|\Psi_v^+\rangle$ and $|\Psi_v^-\rangle$ can be written as linear combinations of the two deperturbed states $|\Psi_1\rangle$ and $|\Psi_2\rangle$

$$|\Psi_v^\pm\rangle = c_1^\pm |\Psi_1\rangle + c_2^\pm |\Psi_2\rangle \,. \tag{1.37}$$

The coefficients $c_1^\pm$ and $c_2^\pm$ for the two wavefunctions can be obtained by variationally minimize the vibrational energy, resulting in the eigenvalue equation

$$\begin{pmatrix} \langle\Psi_1|\mathcal{H}^*_{\text{eff}}|\Psi_1\rangle & \langle\Psi_1|\mathcal{H}^*_{\text{eff}}|\Psi_2\rangle \\ \langle\Psi_2|\mathcal{H}^*_{\text{eff}}|\Psi_1\rangle & \langle\Psi_2|\mathcal{H}^*_{\text{eff}}|\Psi_2\rangle \end{pmatrix} = \begin{pmatrix} \nu_1^* & k \\ k & \nu_2^* \end{pmatrix} \tag{1.38}$$

$$\begin{pmatrix} \nu_1^* & k \\ k & \nu_2^* \end{pmatrix} \begin{pmatrix} c_1^\pm \\ c_2^\pm \end{pmatrix} = \nu_\pm \begin{pmatrix} c_1^\pm \\ c_2^\pm \end{pmatrix} \tag{1.39}$$

where the diagonal matrix elements are the deperturbed energies of the two resonant modes, while the off-diagonal elements are the interaction terms between the two states, which can be estimated by assuming harmonic wavefunctions for $|\Psi_1\rangle$ and $|\Psi_2\rangle$ and a potential of the type of Eq. 1.16 in the Hamiltonian. The eigenvalues of Eq. 1.39 are the energies of the two resonant states and introducing the quantity $\epsilon^* = |\nu_1^* - \nu_2^*|/2$ they can be written as

$$\nu_\pm = \frac{\nu_1^* + \nu_2^*}{2} \pm \sqrt{\epsilon^{*2} + k^2} \tag{1.40}$$

The quantity in the square root is thus half the magnitude of the splitting between the two frequencies induced by the Fermi resonance and in Figure 1.1 it has been plotted as function of k and ϵ^*.

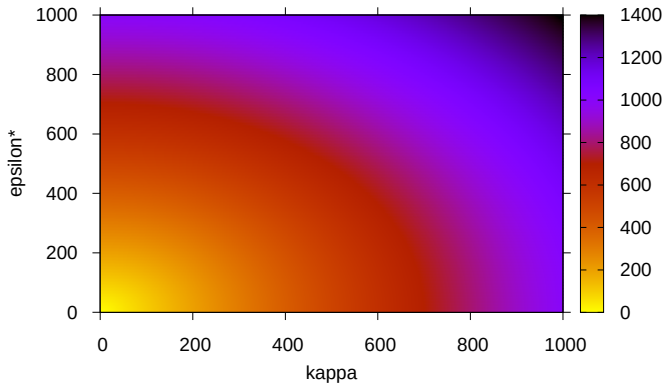


Figure 1.1: Half the magnitude of the frequency splitting due to the Fermi resonance as a function of ϵ^* and k, within the variational (exact) treatment. The maximum value of the function in the represented domain is $\sqrt{2 \cdot 10^6}$ at $\epsilon^* = k = 1000$, on the top right edge of the figure.

In this picture, ϵ^* should be viewed as a parameter rather than a true variable, since we have assumed at the beginning that the two modes were resonant, subsequently applying the variational treatment. Then, Figure 1.1 tells us that given two modes of fixed deperturbed frequency separation ϵ^*, the magnitude of the frequency splitting induced by the Fermi resonance is a linear function of k. So, in the limit situation of $\epsilon^* \to 0$, and $k = 1000$ cm^{-1} the frequency of the two modes will differ by 1000 cm^{-1}.

Moving to the perturbation theory framework, we can consider the off-diagonal terms in Eq. 1.38 as an higher order perturbation with respect to the diagonal ones and the square root of Eq. 1.40 can be expanded in Taylor series up to the second order[g]

$$\nu_{\pm} = \frac{\nu_1^* + \nu_2^*}{2} \pm \left(\epsilon^* + \frac{k^2}{2\epsilon^*} \right) \tag{1.41}$$

The quantity on the right of the $\pm$ sign is then the perturbative expansion of the variational – exact – frequency splitting and it has been reported in Figure 1.2.

[g]The function $\sqrt{1+x^2}$ can be expanded in Taylor series about $x = 0$ as

$$\sqrt{1+x^2} \simeq 1 + \frac{x^2}{2} - \frac{x^4}{8} + ...$$

From a formal point of view an arbitrarly small perturbation parameter can be used, so that $k \to \lambda k$ in Eq. 1.38 and the validity of the expansion is ensured by the fact that λ is arbitarly small.

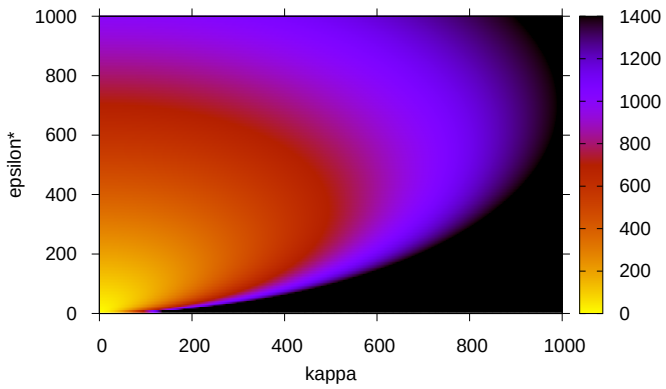


Figure 1.2: Half the magnitude of the frequency splitting due to the Fermi resonance as a function of ϵ^* and k, within the perturbative treatment. The black color has been used for values larger than $\sqrt{2 \cdot 10^6}$.

The same color range has been used in Figure 1.1 and 1.2 and in the latter all the values larger than the maximum of the former have been reported in black. It can be observed that for small values of the ratio k^2/ϵ^*, the perturbative approach is a good approximation of the variational treatment, being the series expansion a good representation of the square root function. However, the condition at which the Fermi resonances usually occur corresponds to large values of k^2/ϵ^*, for which the series expansion is not valid anymore, since the perturbation becomes large and the off-diagonal terms become large corrections to the deperturbed values. The focal point, is that such a behavior is not due to the physical effect, since it does not appear in the variational (exact) treatment of the resonance, but it is rather due to the breakdown of the perturbation theory.[h] As a consequence, in Figure 1.2 it can be observed that due to the presence of ϵ^* at the denominator of Eq. 1.41 a divergence occurs in a large domain (bottom left). When such a divergence occurs, the fourth-order term of the serie in Eq. 1.41 is an upper bound of the error of the second order perturbative treatment, as suggested by Martin *et al.*,[26] so that an estimation of the magnitude of such an error can be given by

$$\frac{k^4}{4\epsilon^{*3}} \tag{1.42}$$

[h] In this conditions the series expansion of the square root function should be about $x \to \infty$

$$\sqrt{1+x^2} \simeq x + \frac{1}{2x} + \frac{1}{8x^3} + ...$$

Let's analyze how the variational and perturbative treatments of the Fermi resonances can be applied to the VPT2 approach introduced in the previous sections. In the VPT2 approach when the harmonic frequencies related to different modes are close to each other, singularities occur in Eq. 1.23, 1.24, 1.25 and Eq. 1.35. Such singularities are usually classified as Fermi type 1 and type 2, corresponding to the conditions $2\omega_i \simeq \omega_j$ and $\omega_i + \omega_j \simeq \omega_k$, respectively. In the expression of the transition dipole moment for the fundamentals (Eq. 1.35) also another type of singularities occurs, when the harmonic frequencies of two fundamental modes are nearly degenerate, ($\omega_i \simeq \omega_j$) and it is commonly refferred to as type 1-1 resonances. Thus, the VPT2 treatment has a behavior very similar to the one showed in Figure 1.2, and the problem which has to be faced is how it is possible to distinguish whether the large terms occurring in the expansion, are due to a large physical splitting of the frequencies induced by the real Fermi resonances of to the divergence due to the breakdown of the perturbation theory.

First of all, it is convenient to expand the potentially resonant terms in partial fraction decomposition

$$\frac{(8\omega_i^2 - 3\omega_j^2)k_{iij}^2}{16\omega_j(4\omega_i^2 - \omega_j^2)} = \frac{k_{iij}^2}{32}\left(\frac{4}{\omega_j} + \frac{1}{2\omega_i + \omega_j} - \frac{1}{2\omega_i - \omega_j}\right) \tag{1.43}$$

$$\frac{\omega_k(\omega_i^2 + \omega_j^2 - \omega_k^2)k_{ijk}^2}{2\Delta_{ijk}} = \frac{k_{ijk}^2}{8}\left(-\frac{1}{\omega_i + \omega_j + \omega_k} + \frac{1}{\omega_i - \omega_j - \omega_k} - \frac{1}{\omega_i - \omega_j + \omega_k} + \frac{1}{\omega_i + \omega_j - \omega_k}\right) \tag{1.44}$$

using the right hand side of Eq. 1.43 and 1.44 in Eq. 1.23, 1.24 and 1.25 in lieu of the original terms, so that in presence of Fermi resonances only the last addenda diverge.

The simplest approach to remove the singularities is to define a threshold value for the harmonic frequency differences, usually of the order of 10 cm^{-1}, evaluate the harmonic frequency differences for all the modes involved in the sums of Eq. 1.23, 1.24, 1.25 and 1.35, and neglect all the contributions in which the difference is smaller than the threshold. The same approach can be also applied to the treatment of the 1-1 resonances occurring in the trasition dipole moment (Eq. 1.35). However, within this approach the effect of the force constant on the overall resonance is neglected, so that a more refined scheme, based on a two-step procedure is often used for the identification of the resonant dyads:

1. The harmonic frequency differences are evaluated for all the possibly resonant modes, and compared with a threshold value, which is usually chosen of 200 cm^{-1}. If the harmonic frequency difference of a given dyad is lower than the threshold, such a dyad is selected as possibly resonant.

2. The quantity[26] in Eq. 1.42 is evaluated for all the possibly resonant terms identified in the previous step. If for a given dyad it is larger than 1, the corresponding contribution is neglected, otherwise the VPT2 treatment is applied without any modification.

Such an approach is usually referred to as the "deperturbed" VPT2 (DVPT2[24–26]). So, the underlying idea of the DVPT2 approach is to neglect the resonant terms on the basis of the identification criteria illustrated above, so that the singularities do not occur. Nevertheless, in this way the term involving the force constants related to the resonant dyads do not give any contribution to the anharmonic frequencies, so that also the physical effect of the Fermi resonances is entirely removed. For this reason a mixed perturbative/variational scheme has been developed, based on the variational treatment of the resonances illustrated above. In particular, an interaction matrix similar to the one presented in Eq. 1.38 is built, using in the diagonal elements the DVPT2 frequencies as an approximation of the deperturbed frequencies. Then the eigenvalues of such a matrix are the variationally corrected anharmonic frequencies of the resonant modes. When more than one resonant dyad occur, a unique interaction matrix including all the resonant states is constructed and by diagonalizing such a matrix all the frequencies are simultaneously obtained. Such a treatment is called the "generalized" VPT2 (GVPT2[24–26]) scheme, and is formally more consistent than the DVPT2 scheme, since the neglected force constants are included in the off-diagonal terms as higher order corrections to the deperturbed energies.

The main drawback of the GVPT2 and DVPT2 treatments is that they are directly threshold dependent, and in some cases this can have unexpected consequences. For example, when the anharmonic frequencies computed using different force fields are compared, it can occur that the harmonic frequency differences and the Martin criteria are close to the threshold values, so that different resonant dyads are identified with the different methods. In this case, the differences between the two sets of anharmonic frequencies are not fully due to the force fields, but also to the identification of the resonant dyads, and it becomes difficult to separate the two effects. Moreover, when the spectra are computed on-the-fly in a molecular dynamics simulation, it can occur that for two subsequent snapshots different resonant dyads are found, leading to awkward steps in the thermodynamical functions. In both these cases, the differences in the identification of the resonant dyads may be affected by the threshold values rather than being due only to physical effects. For these reasons, other schemes have been proposed. One of the most effective ones is based on a variational treatment of the resonant modes, analogous to that showed in Eq. 1.38 and 1.39 but in which the harmonic frequencies are used as the diagonal matrix elements. By analogy with the formalism introduced above, the resonant terms can be written as

$$\frac{Sk^2}{\epsilon} \tag{1.45}$$

where ϵ is defined as half the difference between the harmonic frequencies of the resonant modes, k depends on the force constant coupling the modes, and S is a sign function related to the frequency difference.[25] Then, Eq. 1.41 can be used to

define the transformation

$$\frac{Sk^2}{\epsilon} \quad \rightarrow \quad \sqrt{k^2+\epsilon^2} - S\epsilon \tag{1.46}$$

If all the possibly resonant terms in Eq. 1.23, 1.24 and 1.25 are approximated as in Eq. 1.46, the approach is called "degeneracy-corrected" VPT2 (DCPT2[25,43]), and the singularities never occur whatever the values of k and ϵ. As a consequence, no thresholds are required for the identification of the resonant dyads. Nevertheless, the validity of such an approximation should be carefully evaluated, since it is based on the expansion of the square root function about $k = 0$ and valid only when k is small. In this case the Taylor series is a good approximation of the square root function and the DCPT2 approach is equivalent to VPT2, having also the advantage that when ϵ is small it never diverges, while VPT2 does. As the value of k increases, the Taylor series is not a good approximation of the square root function and DCPT2 becomes less correct. So that when ϵ and k are both large, VPT2 should be used.

Then, when k is small in most cases DCPT2 and the VPT2 approaches can be interchangeably applied, in the limit case in which k and ϵ are both large the VPT2 approach should be used without any correction and it should be preferred, being more correct than the DCPT2. In the limit case of k^2/ϵ large (usually when $\epsilon \rightarrow 0$) only the DCPT2 can be used, since the VPT2 diverges. This trend can be expressed as a function of the product $|k \cdot \epsilon|$ and the resonant terms of Eq. 1.45 can be interpolated using a damping function $f(|k \cdot \epsilon|)$, which smoothly passes from the VPT2 to the DCPT2 form

$$\begin{aligned} &\frac{k^2}{\epsilon} \rightarrow f(\epsilon,k)\frac{k^2}{\epsilon} + [1-f(\epsilon,k)](\sqrt{k^2+\epsilon^2}-\epsilon) \\ &f(\epsilon,k) = \frac{\tanh\left[\alpha\left(\sqrt{\epsilon^2 k^2}-\beta\right)\right]+1}{2} \end{aligned} \tag{1.47}$$

When Eq. 1.47 is applied to all the possibly resonant terms the scheme is called the "hybrid" DCPT2 (HDCPT2[25]) model. Since at $\epsilon = 0$ the damping function tends to one

$$\lim_{\epsilon \rightarrow 0} \frac{\tanh(\epsilon)}{\epsilon} = 1 \tag{1.48}$$

the divergence never occurs in the HDCPT2 approach. The two parameters of the damping function α and β can be tuned in order to increase or decrease the fraction of the VPT2 and DCPT2 contributions and to modify the smoothness of the transition from one model to the other. In the current implementation of the HDCPT2 scheme the values of 1.0 and $5{\cdot}10^5$ have been used for α and β, respectively. In Figure 1.3 $f(\epsilon,k)$ has been plotted as a function of the product $|k \cdot \epsilon|$, in the zone of the shifting from the VPT2 to the DCPT2 model. It can be observed that the current value of β leads to a 50% of mixing at $|k \cdot \epsilon| = 500000$, while α has been chosen in order to produce a step-like function, which quickly passes from one model

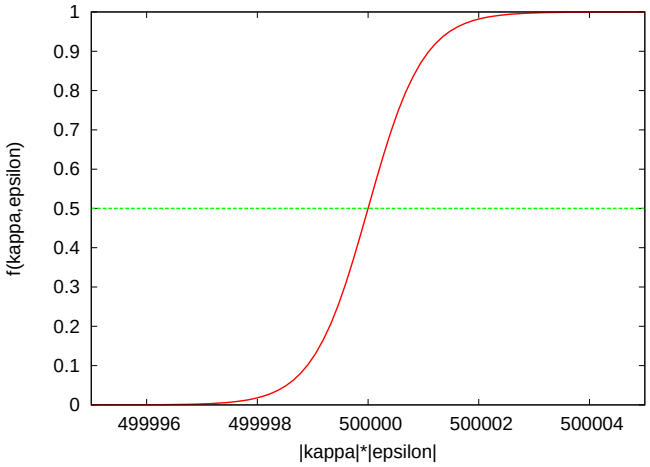


Figure 1.3: Damping function used in the HDCPT2 approach.

to the other. In fact, the VPT2 has a weight lower than 2% for $|k \cdot \epsilon| = 499998$, and larger than 98% for $|k \cdot \epsilon| = 500002$, so that the whole step is mostly included in an energy range of 4 cm^{-1}. The phylosophy behind such a choice is to be sure that the domain in which the DCPT2 model is applied is large enough to exclude the unphysical divergences of the VPT2 model.

1.1.5 The theoretical Infrared spectra

From the transition probability of Eq. 1.8 the *molar absorption coefficient* can be estimated, by summing over all the possible $I \rightarrow F$ transitions and eventually passing from the molecular to the molar quantity[140,148]

$$\varepsilon(\nu) = \frac{4N_A\pi^2}{3 \cdot 10^3 \ln(10)hc}\nu \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \delta(\nu_{IF} - \nu) \tag{1.49}$$

where $\nu_{IF} = \nu_F - \nu_I$, N_I is the Boltzmann population of the state I, $\langle \boldsymbol{\mu} \rangle_{IF}$ is the transition dipole moment and N_A is the Avogadro constant (mol^{-1}). The spectral lineshape is here given by the Dirac delta function since no information about the relaxation dynamics can be obtained at this level of theory.

By integrating the *molar absorption coefficient* in a frequency interval $[\nu_1, \nu_2]$ (see Appendix A), the theoretical estimation of the *integrated cross section* can be obtained

$$G_{\nu_1,\nu_2} = \int_{\nu_1}^{\nu_2} \varepsilon(\nu) d\nu = \sum_{\substack{IF \\ \nu_1 < \nu_{IF} < \nu_2}} G_{IF} \tag{1.50}$$

with

$$G_{IF} = \frac{4N_A\pi^2}{3 \cdot 10^3 \ln(10)hc} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \nu_{IF}. \tag{1.51}$$

Eq. 1.50 suggests that the Infrared spectrum can be simulated as the sum of a certain number of $I \rightarrow F$ transitions, the G_{IF} quantity corresponding to the area subtended the peak of each transition. The evaluation of the G_{IF} requires the knowledge of the transition frequencies ν_{IF} and dipole moments $\langle \boldsymbol{\mu} \rangle_{IF}$, which can be computed at either the harmonic or anharmonic level, as discussed in Section 1.1.3. Once the G_{IF} are known, the G_{ν_1,ν_2} can be calculated and compared with the experimental ones of Eq. 1.3. Such a comparison is straightforward as long as we assume that the experimental cross sections vanish at the boundaries of the integration intervals (i.e., the area associated to a specific transition is completely included within the integration limits). Furthermore, by convoluting the G_{IF} with Gaussian (or Lorentzian) functions, the *absorption cross sections* can be obtained (see Appendix A), allowing a direct comparison between the simulated and the experimental lineshapes. In this picture, the full width at half maximum of either the Gaussian or Lorentzian function is simply an adjustable empirical parameter.

Generally speaking, although the harmonic frequencies and intensities can be useful for the assignment of the peaks of very simple experimental spectra, there is often a specific need to go beyond such an approximation. In particular, when large systems are involved the Infrared spectra appear as a superposition of many peaks, and the positions computed within the harmonic approximation can lead to errors of the order of 100 cm^{-1} for modes with large anharmonicities (e.g., C-H stretching modes). Furthermore, due to the occurrence of the Fermi resonances and of the overtone and combination bands the experimental spectra can be drastically different from the harmonic ones. The anharmonic Infrared spectra, besides including the features originating from such non-fundamental transitions, also provide very accurate fundamental frequencies, as it will be shown at the end of this Chapter. Basically, the accuracy of the results will be dependent on the accuracy of the force constants (ω_i, k_{ijk}, k_{ijkl}) and of the derivatives of the electric dipole moment ($\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\mu}_{ijk}$). Thus, the proper choice of the suitable level of theory for the treatment of the electronic problem is a fundamental step in the calculation of the Infrared spectra. In the next section we will discuss the results which can be expected in these types of calculations.

1.2 Numerical results

In the following three sections, some numerical result will be shown in order to provide a general perspective about the accuracy that could be expected from the VPT2 approach for both the frequencies and the intensities. All the three sections have been taken from works published in the previous years, so that only the relevant aspects will be discussed here and the interested reader is addressed to the original works. In particular, in Section 1.2.1 the Infrared spectrum of the CH_2ClF will be discussed, as computed at the full coupled-cluster level;[73] in Section 1.2.2, results obtained with the Density Functional Theory (DFT) will be discussed for a set of

eight halo-substituted hydrocarbons;[70] and in Section 1.2.3 an application of the VPT2 for the conformational analysis of the gas-phase mixtures will be shown.[71]

1.2.1 The CH_2ClF molecule

One common high level approach for the calculation of vibrational properties is the coupled-cluster method, including the explicit treatment of the single and double excitations, followed by a perturbative treatment of triple excitations – CCSD(T)[149] –. Even when the non Born-Oppenheimer coupling and the explicit treatment of the scalar relativistic effects are neglected, the CCSD(T) method furnishes very accurate results, for both frequency and intensities, once provided a proper choice of the basis set. The major limitation of such a method is the large computational time and memory required by the calculations. Indeed, this type of calculations can only be performed for small systems, being increasingly less feasible when molecules larger than benzene, are involved. In this section we will show the results obtained for one molecule – CH_2ClF – which is small enough to allow the full evaluation of the CCSD(T)/aug-cc-pVTZ (AVTZ) anharmonic force field and anharmonic intensities with the VPT2 approach. The CH_2ClF molecule is quite a representative case, since the presence of halogens makes the calculation of vibrational properties non trivial due to the large electronegativities of the fluorine and chlorine atoms. Being composed by only five atoms, the whole vibrational motion can be described with nine normal modes. For comparison, the same calculations have been performed also at the B2PLYP[67–69] and MP2[65] levels, using the same basis set. The computational time required by the coupled-cluster computations was about two weeks, to be compared with the two days required by DFT and MP2 computations using the same basis set and the same number of processors and nodes. In Table 1.1 the fundamental frequencies computed with all the methods and experimentally measured have been reported. By comparing the harmonic and anharmonic CCSD(T)/AVTZ frequencies

			Harm.	Anharm.		
Symmetry	Mode	Exp.	CCSD(T)/AVTZ	CCSD(T)/AVTZ	B2PLYP/AVTZ	MP2/AVTZ
A'	ν_1	2992.0	3102.2	2989.7	3015.7	3036.4
A'	ν_2	1473.6	1518.8	1472.4	1476.2	1487.5
A'	ν_3	1353.3	1384.1	1349.0	1352.6	1364.2
A'	ν_4	1067.8	1091.2	1065.6	1045.3	1063.4
A'	ν_5	759.9	767.6	748.4	735.1	773.9
A'	ν_6	383.5	385.2	380.2	377.7	385.7
A"	ν_7	3035.4	3178.6	3027.8	3053.2	3080.9
A"	ν_8	1236.8	1261.0	1226.8	1233.4	1241.5
A"	ν_9	1000.8	1018.5	997.7	1000.7	1012.8
MAE			44.9	5.1	11.3	16.9

Table 1.1: Comparison between experimental and computed frequencies (cm^{-1}). Mean Absolute Error (MAE) computed with respect to the experimental frequencies. See Ref.[73] for further details.

it is evident that the contribution from the cubic and quartic force constants strongly

improve the accuracy of the frequencies, the MAE passing from 45 cm^{-1} to 5 cm^{-1}, respectively. In particular, for the accurate evaluation of the frequency of the C-H stretching modes (ν_1 and ν_7) the use of an anhamonic force field improve the results by more than 100 cm^{-1}, leading to a quantitative agreement with the experimental data. The MAE of the anharmonic fundamental frequencies at the B2PLYP and MP2 levels are 11 and 17 cm^{-1}, respectively. In Table 1.2 the experimental and theoretical $G_{\nu_1\nu_2}$ have been reported for the same molecule. In the low frequency

$[\nu_1, \nu_2]$ (cm^{-1})	main transition	Exp.	Harm. CCSD(T)/AVTZ	Anharm. CCSD(T)/AVTZ	Anharm. B2PLYP/AVTZ	Anharm. MP2/AVTZ
950-1175	ν_9,ν_4	144	162.20	156.91	164.80	160.26
1175-1290	ν_8	2.40	2.43	2.37	2.35	2.29
1290-1410	ν_3	25.14	31.74	27.86	26.67	27.08

Table 1.2: Comparison between experimental and computed integrated cross sections (km/mol). Only fundamental modes have been considered here. See Ref.[73] for further details.

zone of the IR spectrum the main contributions to the overall intensities derive from the fundamental modes, since overtones and combination bands occur – by definition – at higher frequencies. Thus, the comparison between experimental and harmonic intensities is meaningfull. From Table 1.2 it is evident the large improvement of the results when passing from the harmonic to the anharmonic treatment of the transition dipole moment, and also in this case the coupled-cluster results have an accuracy higher than the other two methods.

1.2.2 Other halo-hydrocarbons molecules

Methods based on the Density Functional Theory (DFT) are often the methods of choice when larger systems are involved, since they usually provide a good accuracy at a feasible price in terms of computational demands, and they can be applied to small- to medium-sized systems. One of the most reliable methods is given by the combination of the B3LYP[66] functional with double-ζ polarization augmented basis sets, in particular of the SNSD[150] family. Such a basis set is an improved version of the polarized double-ζ N07D basis set,[144,150–153] obtained by adding diffuse s-functions on all the atoms, diffuse polarized d-functions on heavy atoms (p on hydrogens) and Stuttgart-Dresden electron core pseudopotentials[154,155] to bromine and iodine. In fact, although the original N07D provided very good results in the computation of frequencies and EPR properties,[6,51,71,74–77,79,144,150–153] the inclusion of diffuse functions and pseudopotentials are needed to properly treat the heaviest halogen atoms and to improve the performances of the IR intensity calculations. The double-hybrid B2PLYP[67,68] functional, in conjunction with the cc-pVTZ basis set[156–158] (cc-pVTZ-PP[159,160] in case of Br and I) is also a very well stated method, for which analytic second derivatives[69] with respect to the nuclear displacements are available. Although the inclusion of a portion of the MP2[65] energy and the use of a triple-ζ basis set lead to an higher computational cost, such a method

usually provides very accurate harmonic frequencies[69] and may improve anharmonic corrections in problematic cases.[77]

In a recently published work,[70] extensive studies of the performences of such DFT methods in the computation of vibrational properties have been presented for eight halo-substituted hydrocarbons (halo-methanes: CH_2F_2, $CHBrF_2$, CF_3Br, CH_2DBr; halo-ethylenes: CH_2CHF, CF_2CFCl, cis-CHFCHBr, cis-CHFCHI), and the results are summarized in Table 1.3. The overall MAE of the B3LYP/SNSD level

	B3LYP/ SNSD		B2PLYP/ cc-pVTZ(-PP)[a]	
	freq.	intens.	freq.	intens.
CH_2F_2	28.8	3.40	6.9	3.22
$CHBrF_2$	23.1	6.55	6.9	6.18
CF_3Br	26.0	8.21	9.8	9.78
CH_2DBr	9.9	-	7.3	-
CH_2CHF	8.5	11.05	11.6	7.57
CF_2CFCl	12.5	4.82	7.2	3.16
cis-CHFCHBr	8.0	-	8.2	-
cis-CHFCHI	8.0	-	8.0	-
halo-methanes (36)	22.0	6.05	7.7	6.39
halo-ethylenes (48)	9.3	7.94	8.8	5.37
total (84)	14.7	6.81	8.3	5.98
C-H st. (13)	8.8	-	13.9	-
C-F st. (11)	32.0	-	9.2	-
C-Cl st. (1)	8.2	-	1.0	-
C-Br st. (4)	20.4	-	6.3	-
C-I st. (1)	18.6	-	0.4	-

Table 1.3: Mean Absolute Errors for the anharmonic (GVPT2) frequencies (cm^{-1}) of the fundamental modes and *integrated absorption cross sections* (km/mol), with respect to the experimental data. Anharmonic intensities of the overtones and combination bands have also been included. See the original work[70] for all the details about each molecule. The number of stretching modes considered over the whole set of molecules is given in parenthesis.

is about 15 cm^{-1}, which is comparable with previous results, obtained for example for glycine[71] and uracil[6] (MAEs of about 10 cm^{-1} in both cases). The slightly larger errors found in the present case are related to the presence of the halogen atoms, which introduce further degrees of complexity due to the large electronegativity and – for the heavy-atom-containing compounds – to the relativistic effects. In particular, the frequencies of the C-F, C-Br and C-I stretchings are affected by the largest errors (about 30 cm^{-1} for C-F and 20 cm^{-1} for C-Br and C-I), which are the origin of the large MAEs observed for CH_2F_2 (29 cm^{-1}), $CHBrF_2$ (23 cm^{-1}) and CF_3Br (26 cm^{-1}). The errors on the frequencies for modes associated to the chlorine atom are rather limited, being only 8 cm^{-1} for the C-Cl stretching mode of CF_2CFCl. Overall better performances are observed in B2PLYP/cc-pVTZ(-PP) computations, with a mean error on the frequencies of the C-F stretching modes of only 9 cm^{-1}, and MAEs of about 1 cm^{-1} for C-Cl and C-I and 6 cm^{-1} for the

C-Br stretchings. On the whole, this leads to overall MAEs smaller than 10 cm^{-1} for the fundamental frequencies of CH_2F_2, $CHBrF_2$ and CF_3Br. Therefore, the use of a triple-ζ basis set and the inclusion of a fraction of the MP2 energy leads to a noticeable improvement of the anharmonic frequencies with respect to the B3LYP/SNSD results, thus suggesting that the polarization effects induced by the large electronegativity of the fluorine atoms can be accounted for with a triple-ζ basis set. Nevertheless, the B3LYP/SNSD approach remains a valuable alternative for open-shell systems or for larger molecules, for which the B2PLYP/cc-pVTZ method is much more prone to spin contamination errors and/or much more computationally demanding.

For CH_2F_2, $CHBrF_2$, CF_3Br CH_2CHF and CF_2CFCl the experimental *integrated absorption cross sections* are available, and in Table 1.3 a summary of the MAEs of the computed $G_{\nu_1\nu_2}$ with respect to experimental ones is given. We observe that for both the proposed DFT methods the agreement between theoretical and experimental intensities is very good, the total MAE being about 7 km/mol for B3LYP/SNSD and 6 km/mol for B2PLYP/cc-pVTZ(-PP). On average, both methods provide very similar intensity values, thus suggesting that electrostatic properties should also be in close agreement and that the effect of the double-/triple-ζ basis sets should be smaller than that observed for frequencies. Furthermore, a similar accuracy is found for both halo-methanes (MAEs of about 6 km/mol for both B3LYP/SNSD and B2PLYP/cc-pVTZ(-PP)) and halo-ethylenes (MAE of about 8 km/mol for B3LYP/SNSD and about 5 km/mol for B2PLYP/cc-pVTZ(-PP)) because infrared intensities are less sensitive to the molecular geometry than frequencies, being lower-order derivatives with respect to atomic displacements, thus suggesting that the performances of the methods should be quite similar for a large set of molecular systems.

In Figure 1.4 the IR spectrum of cis-CHFCHI has been shown, as computed at B3PLYP/SNSD and B2PLYP/cc-pVTZ-PP levels and compared with the experimental one.

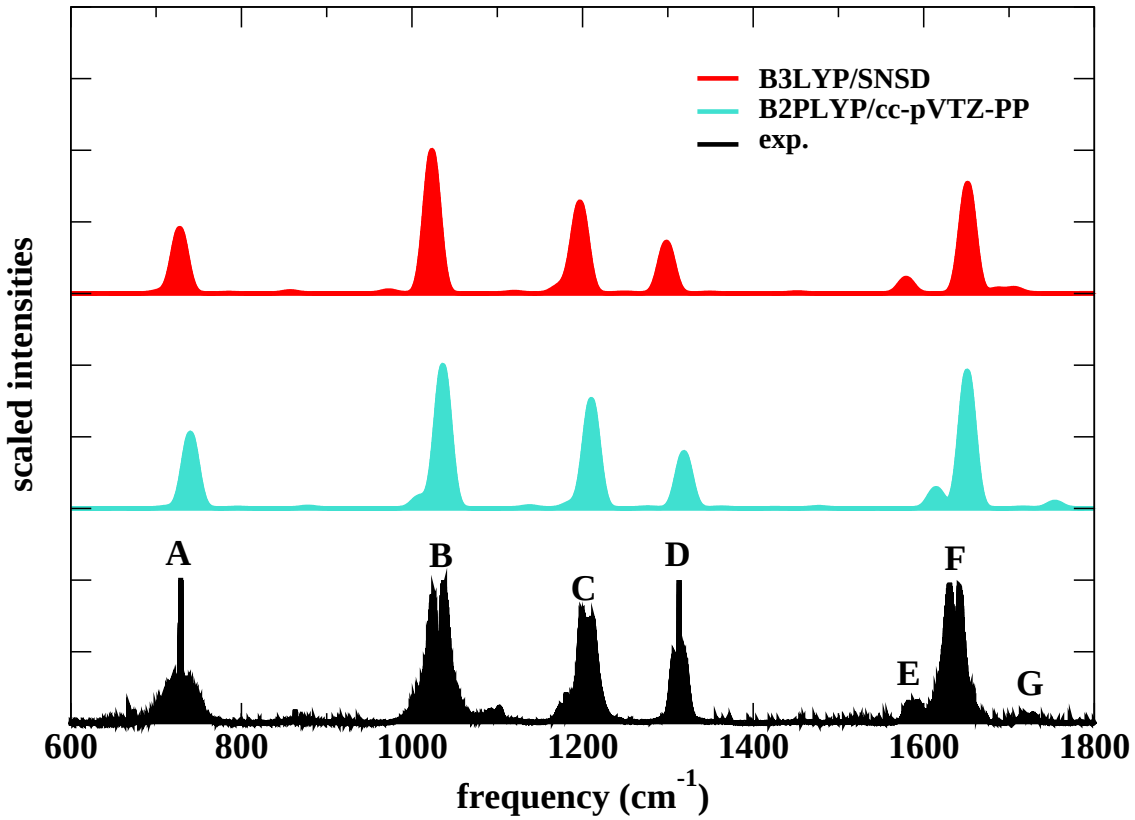


Figure 1.4: Infrared spectrum of cis-CHFCHI. A full width at half maximum of 20 cm^{-1} has been used for the convolution. [A] CCF and CHI out of plane bending. [B] CF stretching. [C] CHI/CHF symmetric in plane bending. [D] CHI/CHF asymmetric in plane bending. [E] Overtones and combination bands of CCF bending. [F] CC stretching. [G] Overtones and combination bands of CHI out of plane bendings.

1.2.3 Test case: Gas-phase glycine conformers.

Glycine is the simplest natural amino-acid and it has been extensively studied both experimentally[4,5,161–167] and theoretically.[4,5,18,168–174] Recently, glycine has been the object of particular attention due to its possible presence in interstellar space and to its use in the design of bio-nanodevices.[20,51,175–177] Additionally, its conformational flexibility due to the presence of three internal rotational degrees of freedom leads to several possible local minima separated by low energy barriers. This rather complex PES makes glycine an interesting specimen for computational methodologies. In fact, it has become a test case for the studies of conformational stability (for a detailed list of works before 1992 see Ref.[170,171,178]). Nowadays, there is a general agreement on the structure and energetics of the two most stable conformers.[18,168–172,179] Lately, another slightly less stable conformer has received particular attention from theoretical[18,168,169,172,180] and experimental[5,163] points of view. Extensive works on the thermodynamical stability and vibrational properties of the gas-phase conformers of the glycine molecule have been performed also in our group in the past years.[71,75–77,168]

Even at low temperatures, glycine in the gas phase consists of a mixture of at least three conformers, shown in Figure 1.5, and processes of interconversion between them are very likely to take place.[77,181] The most stable conformer of the gas-phase glycine molecule, GLYIp,[i] has a planar geometry, belonging to the C_s symmetry point group, and its structure is stabilized by intra-molecular hydrogen bonds between the -NH_2 and -O(=C) moieties. A slightly less stable local minimum corresponds to the GLYIIn conformer, whose C_s structure is a saddle point,[77,171] while, although the exact structure of GLYIII is still a matter of debate due to the very small energy differences between C_s and C_1 conformers, from the studies performed until now the planar one should be the most stable. [j] In Table 1.4 the relative energies of the three conformers studied in the present work with the B3LYP/aug-N07D[144,151–153] method are reported together with those obtained in other works.[18,77,173]

The relative abundances of such conformers in a gas-phase mixture have been estimated by comparison between experimental IR spectrum and theoretical calcu-

[i] In the notation employed here, "p" stands for "planar" and "n" stands for "non planar"

[j] Csaszar[170] predicted a C_s conformer with an energy of only 0.58 kJ/mol (48.5 cm^{-1}) higher than the C_1 one at the MP2/6-311++G** level of theory. Nevertheless, in that work it was also shown that this value strongly depends on the method and the basis set employed. In fact, using MP2 with a different basis set (see the original work for the details), the stability is inverted, the relative energies becoming 6.98 kJ/mol (583.1 cm^{-1}) for C_s and 7.01 kJ/mol (586.2 cm^{-1}) for C_1, and an extrapolation procedure led to values of 6.67 kJ/mol (557 cm^{-1}) for the C_s and 6.70 kJ/mol (560 cm^{-1}) for the C_1 conformer. According to calculations carried out at the DFT level[168] the GLYIIIp conformer is 0.11 kJ/mol more stable than its C_1 counterpart, yielding also one imaginary frequency for the latter. This is in line with a more recent work,[77] in which the GLYIIIp conformer was found to be a local minimum, with a relative energy of 7.44 kJ/mol using a coupled-cluster force field.

	B3LYP/ aug-N07D	B2PLYP/ aug-cc-pVTZ	MP2/ best estimate[a]	CCSD(T)/ CBS[b]	CCSD(T) DFT[c]	CCSD(T)/ CBS+CV[d]
GLYIp (C_s)	0.0	0.0	0.0	0.0	0.0	0.0
GLYIIn (C_1)	2.0 (3.0)	2.5 (3.5)	2.1 (3.0)	2.7 (3.6)	2.45	2.29 (3.82)
GLYIIIp (C_s)	6.8 (6.7)	6.8 (6.8)	6.7 (6.6)	7.4 (7.3)	7.42	7.44 (7.48)

Table 1.4: Relative energies (in kJ/mol) of the three most stable conformers of glycine. Energies corrected by the anharmonic zero point vibrational energy (ZPVE) are given in parenthesis. [a] Csaszar *et al.*;[170] [b] Balabin *et al.*;[173] [c] CCSD(T)/CBS+CV energies on B3LYP/SNSD geometries;[77] [d] Barone *et al.*.[77]

lations.[4,5] The works of Polfer *et al.*[4] and Adamowicz *et al.*[5] show that a single conformer is predominant with relative abundances of 84% and about 70%, respectively. Qualitatively similar conclusions were reached in an older study employing GED (Gas phase Electron Diffraction) and MW (Micro-Wave) experimental data with the support of a theoretical force field.[162] This experimental evidence matches very well the theoretical predictions, since many different computational methodologies[18,168–172,179] agree on the relative energies of the most stable conformers of the glycine molecule.

In Table 1.5 the harmonic and anharmonic frequencies and intensities computed at the B3LYP/aug-N07D level for the GLYIp conformer are reported. The experimental data[5,163] and the cc-VSCF results obtained at the MP2/DZP level are also shown. A MAE of 9 cm^{-1} has been found when the GVPT2 treatment of the Fermi resonances was employed, the accuracy of the VPT2 model being much higher than the cc-VSCF method which shows a MAE of 35 cm^{-1}. Similar calculations have been done for all the three conformers, and we address the interested reader to the original work.[71] We only recall here that using the GVPT2 model with the B3LYP/aug-N07D force field, the MAE obtained for the frequencies of GLYIIn and GLYIIIp were 20 cm^{-1} and 15 cm^{-1}, respectively.

Such an high accuracy for the frequencies, allowed us to simulate the IR spectrum of the gas-phase mixture using the IR intensities reported in Table 1.5 for GLYIp, and the other data reported in the original work[71] for the other two conformers. The relative abundances (GLYIp=84%, GLYIIn=8%, GLYIIIp=8%) suggested by Polfer *et. al.*[4] were used in order to scale the idividual spectra of the conformers. The IR spectra computed at the VPT2 level and the experimental counterparts measured in low-temperature Ar matrix[5] and in Helium nanodroplets[163] are shown in Figures 1.6 and 1.7, for the 400-2000 cm^{-1} and 2700-3600 cm^{-1} energy ranges, respectively. Also the cc-VSCF spectra obtained from the data reported by Gerber *et. al*[18] (and shown in Table 1.5) is presented. It should be noted that since for the O-H stretching vibration, the most accurate experimental results have been obtained employing the molecular beam depletion spectroscopy,[163] we included the spectrum obtained from such data in the inset of Figure 1.7. In general, similar results can be observed between the VPT2 ad the VSCF spectra for most of the transitions. The most significant discrepancies between the two methods, are in

the spectral range 550-950 cm^{-1}, and from both the graphical representation in Figure 1.6 and from numerical data in Table 1.5 an higher accuracy can be observed when the VPT2 method in conjunction with the B3LYP/aug-N07D force field is used, in particular in the prediction of the band positions and intensities for the features around 900 cm^{-1}. From one side, this can be explained with the larger basis set used for the VPT2 computations, which includes also the diffuse functions, usually important for the intensities. On the other side, this highlights an important drawback of the application of the VSCF approach to large systems. In fact, due to the computational burden related to the PES exploration, such calculations usually must be performed using approximate electronic structure methods.

In Figure 1.8 the Infrared spectra of the individual conformers and of the mixture, computed with the VPT2 approach are reported, and compared with the experimental one in the 550-950 cm^{-1} energy range. The contributions from the different conformers are then evident, in particular the large contribution from the GLYIp one is shown. Also overtones and combination bands occur in such an energy range.

On these ground, the reliability of the VPT2 computations employing basis sets augmented by diffuse functions in the reliable prediction of Infrared spectra can be stated.

		B3LYP/N07D				MP2/DZP[a]		Exp	Assignment
Mode	Symm.	ω	I_{harm}	ν_{GVPT2} (ν_{DVPT2})	I_{anh}	$\nu_{cc-VSCF}$	I_{anh}	Ar matrix[b]	
1	A'	3750	59	3568	47	3598	109	3560 (3585)[c]	OH stretch
2	A"	3582	6	3407	3	3382	5	3410	NH_2 (A) stretch
3	A'	3509	2	3387* (3353)	1	3343	3	-	NH_2 (S) stretch
4	A"	3079	6	2929* (2932)	7	2986	17	-	CH_2 (A) stretch
5	A'	3044	18	2938* (2913)	17	2959	22	2958	CH_2 (S) stretch
6	A'	1816	301	1784	235	1805	414	1779	C=O stretch
7	A'	1673	18	1627* (1651)	19	1669	23	1630	NH_2 bend
8	A'	1453	16	1431	4	1473	13	1429	CH_2 bend x
9	A'	1396	15	1354* (1357)	11	1410	30	1373	CH_2 bend z
10	A"	1382	0	1344* (1345)	0	1377	0	-	CH_2 NH_2 twist
11	A'	1305	14	1290* (1265)	16	1290	23	-	(OH+CH_2x) bend
12	A"	1181	1	1155	1	1185	1	-	CH_2 NH_2 twist
13	A'	1166	118	1123* (1131)	83	1167	99	1136	CN str. + OH bend
14	A'	1129	186	1097	183	1122	232	1101	CO str. + OH bend
15	A'	921	116	870	81	943	3	883	CC str. + NH_2 bend
16	A"	914	3	909	3	970	184	907	CH_2 NH_2 twist
17	A'	821	76	807	100	847	60	801	CC stretch
18	A"	649	90	626	83	613	143	619	OH oop bend
19	A'	634	6	630	5	633	16	-	(NH_2 + OCO) bend
20	A"	507	30	497	26	514	23	500	OH oop bend
21	A'	464	30	461	28	463	38	463 (458.1)[d]	-
22	A'	257	10	261* (262)	10	270	11	- (250.1)[d]	-
23	A"	210	44	230	42	352	54	- (203.7)[d]	-
MAE =				9 (11)		35			
MIN =				2 (2)		0			
MAX =				26 (45)		148			

Table 1.5: Frequencies in cm^{-1} and intensities in km/mol of GLYIp (C_s). GVPT2 refers to the Generalized VPT2 model;[24,25] for modes involved in Fermi resonances (marked with asterisk) the DVPT2 values are also reported in parenthesis. (A) and (S) refer to "asymmetric" and "symmetric" respectively. "Twist" has been used as shortcut for indicating the synchronous bendings of -CH_2- and -NH_2 moieties. Mean Absolute Error (MAE), MIN and MAX errors have been calculated with respect to the experimental data of Ref.[5] for all the modes but mode 1, for which the data of Ref.[163] have been used. [a] Gerber *et al.*;[18] [b] Adamowicz *et al.*;[5] [c] Krasnokutski *et al.*;[163] [d] Raman spectra from Balabin.[161]

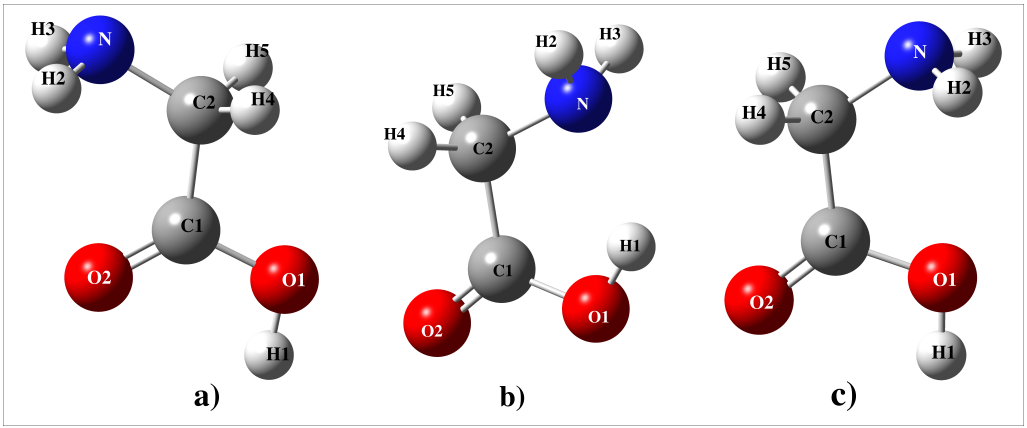


Figure 1.5: The lowest energy conformers of the gas-phase glycine molecule. a) GLYIp(C_s); b) GLYIIn (C_1); c) GLYIIIp (C_s).

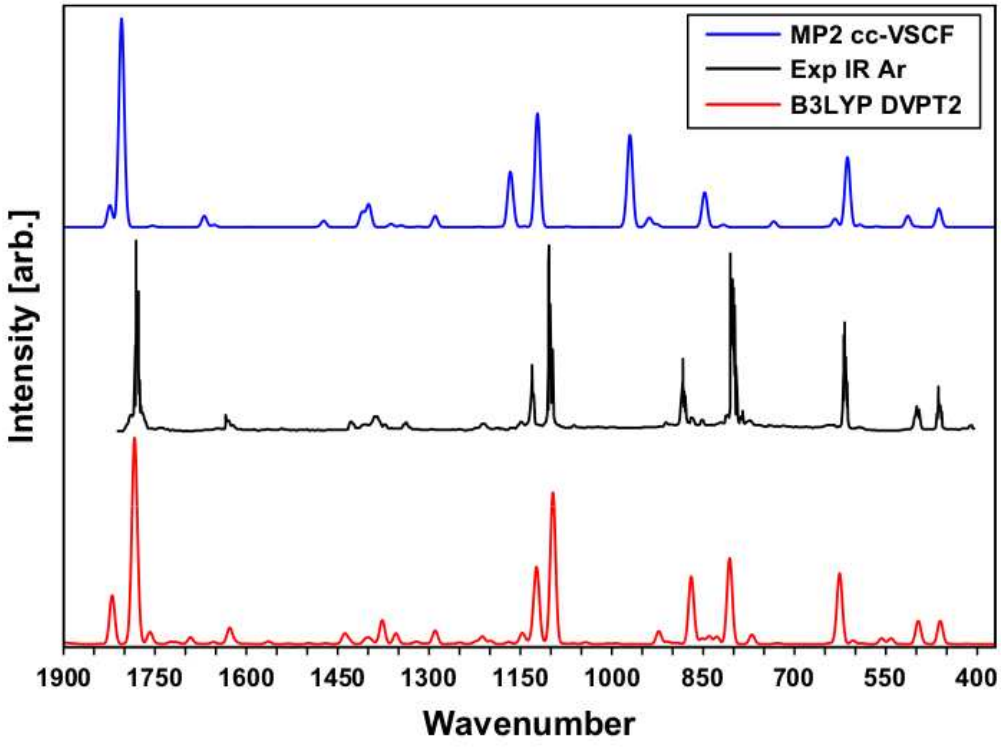


Figure 1.6: Theoretical spectra obtained by summing individual spectra for the three most stable conformers (see text for details) are compared to experiment,[5] in the 400-2000 cm^{-1} energy range.

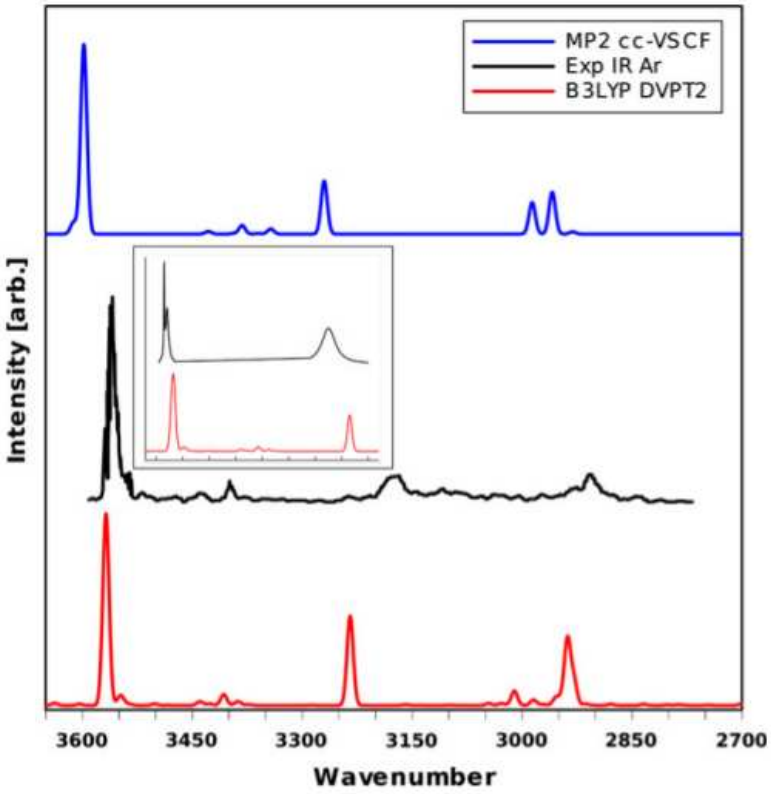


Figure 1.7: Theoretical spectra obtained by summing individual spectra for the three most stable conformers (see text for details) are compared to experiment. Main panel reports the low-temperature Ar matrix IR spectrum[5] in the 2700-3600 cm^{-1} energy range, while the 3200-3600 cm^{-1} spectrum of glycine trapped in helium clusters[163] is shown in the inset.

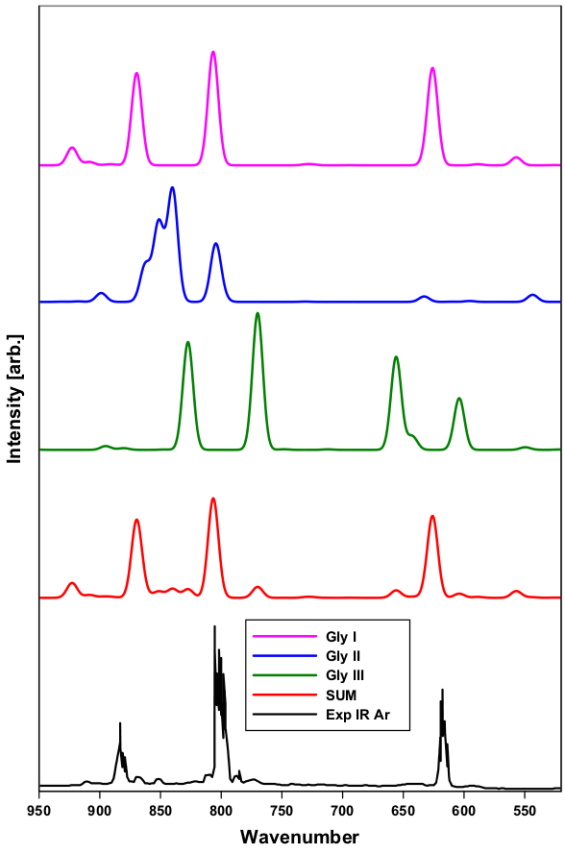


Figure 1.8: Theoretical spectra of the three most stable glycine conformers, along with their sum weighted by relative abundances (see text for details) are compared to experiment[5] for the 550-950 cm^{-1} energy range.

Chapter 2

Hybrid approaches

In the previous Chapter it was shown that when high level methods are used for the computation of the electronic structure, force constants and dipole moment derivatives, the Infrared spectra obtained with the VPT2 approach are in quantitative agreement with the experimental ones. On the other side, less expensive methods lead to results of lower accuracy, e.g. in Section 1.2.1 was reported that for the CH_2ClF molecule, the MAE of the CCSD(T)/AVTZ frequencies with respect to the experimental ones was 5 cm^{-1}, while at the B2PLYP and MP2 levels the MAE was larger than 10 cm^{-1}. One possible way to find a good compromise between the accuracy of the results and the feasibility of the calculations is to combine two different methods for the calculation of the force field. In this Chapter the main aspects of the so-called *hybrid* approaches, deveped in order to improve the accuracy of the calculations for medium/small systems, are first discussed, then, some numerical results obtained with the hybrid methods are presented.

2.1 Theory

2.1.1 The hybrid approach for the frequencies

The frequencies of the fundamental transitions, within the the VPT2 approach can be written as

$$\nu_{0,1_k} = \omega_k + 2\chi_{kk} + \frac{1}{2}\sum_{i\neq k}\chi_{ik}, \tag{2.1}$$

where χ_{ik} are the elements of the anharmonic matrix, $\boldsymbol{\chi}(\mathbf{K},\boldsymbol{\omega})$, depending on the set $\mathbf{K}$ of all the k_{ijk}, k_{ijkl} and Coriolis coupling constants, and on the set $\boldsymbol{\omega}$ of the harmonic frequencies (see Eqs 1.23, 1.24 and 1.25). The anharmonic frequency of mode k can be viewed as the sum of the harmonic frequency and terms depending on the anharmonic matrix, so that Eq. 2.1 can be rewritten in the form

$$\nu_{0,1_k} = \omega_k + \Delta\nu_k(\mathbf{K},\boldsymbol{\omega}). \tag{2.2}$$

$\Delta\nu_k(\mathbf{K},\boldsymbol{\omega})$ is the so called *anharmonic shift* of the fundamental frequency of mode k, the computationally most demanding part of the calculations, requiring the evaluation of the numerical third and fourth derivatives of the PES and the treatment of the Fermi resonances. Usually, the anharmonic shifts are small corrections with respect to the harmonic frequencies, and they are very similar when computed with different methods. As an example, the anharmonic shifts of the C-H stretching mode of $CHBrF_2$ are 4.4%, 4.1% and 4.2%, of the value of the harmonic frequencies at the B3LYP/SNSD, B2PLYP/cc-pVTZ-PP and CCSD(T)/F-AVTZ levels, respectively.[70] Similarly, the anharmonic shift of the C-Cl stretching mode of CF_2CFCl are 1.6%, 1.5% and 1.7% of the harmonic frequencies when the B3LYP/SNSD, B2PLYP/cc-pVTZ-PP and coupled-cluster methods are used.[70] Also for a complex chemical system such as the glycine molecule adsorbed on the (100) surface of Silicon,[51] the anharmonic contributions to the frequencies are of about $3-5\%$ for the high frequency stretching modes, and lower than 3% for all the other modes, and similar percentages are found also when the fully deuterated glycine molecule is considered. Similar results have been reported in many other cases,[6,51,70,71,73,132] and in the next Section the anharmonic shifts of the CH_2ClF molecule computed with different methods[73] will be shown. On the other side, such small corrections are very important when the Infrared spectra need to be computed with an high accuracy, since, say, for a C-H stretching mode typically at about 3000 cm^{-1}, an anharmonic correction of 4% corresponds to 120 cm^{-1}. Thus, when the computation of the anharmonic shift is too demanding, a possible route to decrease the computing time and memory retaining an high accuracy is to use two different methods for the evaluation of the harmonic frequencies and the anharmonic shifts.

As a first approximation, one can run two separate calculations, one for ω_k using a very accurate method, and one for $\Delta\nu_k$ using an approximate method. Then, the two parts can be combined *a posteriori* for the frequencies under analysis. Such an approach will be referred to as HYB in the following. Although this is a very simple and user friendly procedure, not requiring any code implementation, it can lead to inaccuracies due to the fact that the set of $\boldsymbol{\omega}$ used in Eq. 2.2 for the anharmonic shift is computed with a different and less accurate method than the method used for the harmonic part of $\nu_{0,1_k}$. Furthermore, for the modes involved in the Fermi resonances, if $\Delta\nu_k$ is computed with the DVPT2[24–26] or the GVPT2[24–26] schemes, the set of $\boldsymbol{\omega}$ is also empoloyed by the Martin criteria[26] for the identification of the resonant terms, and significant errors due to a different identification of the resonant dyads can occur.[a] In order to overcome such problems, a more refined hybrid approach can be employed, including the harmonic frequencies computed at the higher level of theory directly into the VPT2 treatment, so that the ω_k and the set of $\boldsymbol{\omega}$ appearing in Eq. 2.2 are the same. On the basis of the scheme employed for the treatment of

[a] Also when schemes based on the partial fraction decomposition[43] are employed (such as the DCPT2[25,43] and the HDCPT2[25]) and the identification of the resonances is not a required step, the set of $\boldsymbol{\omega}$ is still necessary for the evaluation of the anharmonic matrix.

the Fermi resonances, we will refer this procedure to as the GHYB approach (when GVPT2 is used for $\Delta\nu_k$) or as the HDHYB approach (when HDCPT2 is used). As a general remark, the GHYB and HDHYB approaches have many advantages with respect to the HYB one. In fact, besides of the higher accuracy and of the more consistent treatment of the Fermi resonances, they also automatically apply the hybrid model to all the frequencies of overtones and combination bands, which otherwise should be computed one by one within the HYB method.

2.1.2 The hybrid approach for the intensities

Regarding to the intensities, the transition dipole moment for a generic fundamental transition (Eq. 1.35) can be splitted in a *doubly harmonic* term (i.e., depending only on $\boldsymbol{\mu}_i$ and $\boldsymbol{\omega}$), and an anharmonic shift (depending also on $\boldsymbol{\mu}_{ij}$, $\boldsymbol{\mu}_{ijk}$ and $\mathbf{K}$),

$$\langle \mu^a \rangle_{0,1_i} = \frac{\mu_i^a}{\sqrt{2}} + \Delta\mu^a(\boldsymbol{\omega}, \mathbf{K}, \mu_i^a, \mu_{ij}^a, \mu_{ijk}^a). \tag{2.3}$$

Using this definition of the transition dipole moments, the G_{IF} of Eq. 1.51 reads

$$G_{0,1_i} = \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10) hc} N_0 \left[\frac{(\mu_i^a)^2}{2} + \frac{\mu_i^a}{\sqrt{2}}\Delta\mu^a + (\Delta\mu^a)^2 \right] \nu_{0,1_i}. \tag{2.4}$$

Hybrid approaches similar to the ones discussed for the frequencies can be applied to the transition dipole moment, by computing the doubly-harmonic part μ_i^a with an high level accurate method while using a cheaper method for the other two terms of Eq. 2.4, requiring the high order numerical derivatives of the PES and of the dipole moment. In practice, this can be done by applying the hybrid approach directly on the *integrated cross sections* as an *a posteriori* correction, in a similar way as the HYB method for the frequencies. It is worth noting that when the GHYB or HDHYB approaches are applied for the frequencies, the set of $\boldsymbol{\omega}$ computed at the higher level is employed also for the calculation of $\Delta\mu^a$, so that the same Fermi resonances are identified. The results from many studies[70,71,73–77] point out the reliability of such an approach, which is thus useful in order to improve the accuracy of the computed intensities.

2.2 Numerical results

The hybrid approaches have been extensively used in the past years,[6,70,71,73–77,132] due to the reliability of the results and of the relatively simplicity of usage, when compared for example to the much more complex reduced dimensionality approach (see next Chapter). In the next sections some selected numerical results will be presented, in order to show the performances which should be expected using such approaches. In particular, in Section 2.2.1 the hybrid frequencies and intensities of the CH_2ClF molecule[73] are shown as computed by a combination of different

CC methods (CC/CC'), of CC and MP2 methods (CC/MP2) and of CC and DFT methods (CC/DFT). Indeed, such a molecule is small enough to allow the harmonic computations at the full CCSD(T) level in the limit of the complete basis set estrapolation. Then, in Section 2.2.2 an overview of the results obtained for the glycine conformers[71,77] and for a set of eight halo-substituted hydrocarbons[70] is shown, in order to present another type of hybrid approaches, which employs two different DFT methods (DFT/DFT') for harmonic and anharmonic computations and is suitable for the treatment of larger systems. The hybrid approach will be also used in Chapter 5, where a DFT/DFTB scheme will be employed for the calculation of the anharmonic frequencies of large systems. In every case, only the concepts useful to the present discussion will be illustrated, while we refer to the original works[70,71,73,77,132] for all the other details.

2.2.1 The hybrid CC/CC', CC/MP2 and CC/DFT models

In Chapter 1 some illustrative calculations on the CH_2ClF molecule were presented. In particular, the anharmonic CCSD(T)/AVTZ force field was found to provide frequecies with an accuracy of about 5 cm^{-1} with respect to the experimental measurements, while at the B2PYLP/AVTZ and MP2/AVTZ levels the MAE were 11 cm^{-1} and 17 cm^{-1}, respectively. The frequencies obtained with such force fields are corrected using the HYB, GHYB and HDHYB approaches, the harmonic part being at the coupled-cluster level, with complete basis set estrapolation and corrections from core-valence interaction and diffuse functions (CCSD(T)/CBS+CV+aug), and the hybrid frequencies are reported in Table 2.1.

As a general comment, we observe that the hybrid CC/CC' model does not provide significant improvements to the CC' frequencies, being the MAE of about 5 cm^{-1} in any case, and this is due to the fact that the harmonic frequencies at the CCSD(T)/AVTZ level are already very accurate, and very small corrections are provided by the correction at the CCSD(T)/CBS+CV+aug level. On the other side, for the proposed MP2 and the DFT methods, the MAE is significantly reduced when passing from the standard GVPT2 to the hybrid models. Apart from the improvement of the agreement with the experimental data, the most relevant point is the treatment of the Fermi resonances and the effects of the force constants on the results, within the different hybrid approaches. So, in the following we will mainly deal with these aspects, addressing to the original work[73] the reader interested in other details of the spectrum of the CH_2ClF.

The identification of the Fermi resonances. For the DFT and MP2 methods, the threshold criteria[26] of the GVPT2 calculations, lead to the identification of one Fermi resonance, the $2\nu_2/\nu 1$. When the GHYB framework is applied, for both the CC/MP2 and CC/DFT force fields two type-1 Fermi resonances ($2\omega_i \simeq \omega_j$) are found, the $2\nu_2/\nu 1$ and $2\nu_6/\nu_5$, related to the CH_2 asymmetric stretching and bending modes and to the CCl stretching and CFCl bending modes, respectively. When the MP2 method is used, for both the resonant modes shown in Table – ν_1

and ν_5 – an error of about 9 cm^{-1} is found for ν_5 with the HYB model, which it is reduced to about 3 cm^{-1} when the GHYB model is applied. This can be due to the fact that when the HYB model is applied, the terms related to the $2\nu_6/\nu_5$ dyad have been included in the anharmonic shift, while within the GHYB model, such terms have been removed. A similar trend is found for the error on another resonant mode – ν_1 –, which passes from 18 cm^{-1} with the HYB model to 12 cm^{-1} with the GHYB model. The errors on the frequencies of these two modes are the main reason behind the higher MAE of the HYB model (7 cm^{-1}) with respect to the GHYB one (5 cm^{-1}), with the MP2 method.

The effect of the force constants. When the GHYB approach is used, the fundamental frequency of ν_1, is in remarkable agreement (within 3.5 cm^{-1}) with the experimental value when the CC/CC' force field is used, while with the CC/DFT and CC/MP2 force fields, it is overestimated by more than 10 cm^{-1}, with the values following the trend CC/DFT > CC/MP2 > CC/CC'. Since the same set of harmonic frequencies is used for all the hybrid models, these differences must be described in terms of cubic and quartic force constants, computed with the different force fields. The values of the quartic force constants obtained with the three methods are quite similar, so that the differences between the CC/CC' and CC/DFT frequencies are mainly due to the terms involving the cubic force constants. Differences larger than 10 cm^{-1} between the DFT and the CC' cubic force constants are found for K_{544}, K_{937}, K_{333}, K_{827}, K_{991} and K_{221}. Among these constants, K_{221} is probably the most important, since it couples mode 1 with its counterpart of the resonant dyad (mode 2), and, when the GVPT2 scheme is applied, it is involved in the off diagonal terms – i.e., the coupling between ν_1 and $2\nu_2$ – of the matrix which is diagonalized. At the DFT and CC' levels the K_{221} values are 184 and 172 cm^{-1}, respectively. This could explain the difference of 13 cm^{-1} between the CC/DFT and the CC/CC' fundamental frequencies of mode ν_1. Within the HDCPT2 approach, the resonant terms are approximated, so that the coupling between the resonant modes is somewhat screened. This could be the reason why the difference between the CC/DFT and CC/CC' frequencies of mode ν_1 is reduced to about 8 cm^{-1}. Similar arguments can be used to discuss the dyad of resonant frequencies $2\nu_6/\nu_5$ and the MP2 results.

To sum up, considering only the fundamentals, the CC/CC' approach reaches an extremely good accuracy, as pointed out by the MAE of about 5 cm^{-1}, while the other two models, CC/DFT and CC/MP2, led to slighly less accurate results but with a much reduced computational cost.

Intensities. The hybrid *integrated cross sections* are computed using with the CC/DFT and CC/MP2 approaches by following the procedure outlined in Chapter 1 and in other works,[1–3,70] and the data are reported in Table 2.2. In the region 700 – 3115 cm^{-1}, which is mainly characterized by absorptions due to fundamentals, the discrepancies between experimental and theoretical estimations are about 9%, while considering the whole spectral range (700 – 6100 cm^{-1}), the MAEs of the two models are 1.93 and 1.95 km/mol, respectively. The two hybrid models perform

equally well in calculating the anharmonic intensities, also in the spectral intervals fully related to the non-fundamental transitions.

			CC	CC'		CC/CC'			shift	
Symmetry	Mode	Exp.	Harm.	GVPT2	HYB	GHYB	HDHYB	HYB	GHYB	HDHYB
A'	ν_1	2992.0	3111.1	2989.7	2998.6	2996.1	2984.1	-112.5	-115.0	-127.0
A'	ν_2	1473.6	1512.8	1472.4	1466.4	1465.1	1461.5	-46.4	-47.7	-51.3
A'	ν_3	1353.3	1379.3	1349.0	1344.2	1348.2	1347.3	-35.1	-31.1	-32.0
A'	ν_4	1067.8	1087.5	1065.6	1061.9	1061.7	1061.7	-25.6	-25.8	-25.8
A'	ν_5	759.9	781.2	748.4	762.0	763.5	769.0	-19.2	-17.7	-12.2
A'	ν_6	383.5	388.3	380.2	383.3	383.9	383.2	-5.0	-4.4	-5.1
A"	ν_7	3035.4	3190.1	3027.8	3039.3	3037.7	3036.6	-150.8	-152.4	-153.5
A"	ν_8	1236.8	1264.1	1226.8	1229.9	1230.3	1230.1	-34.2	-33.8	-34.0
A"	ν_9	1000.8	1022.4	997.7	1001.6	1001.1	1000.9	-20.8	-21.3	-21.5
MAE				5.1	4.7	4.1	5.6			
			CC	MP2		CC/MP2			shift	
Symmetry	Mode	Exp.	Harm.	GVPT2	HYB	GHYB	HDHYB	HYB	GHYB	HDHYB
A'	ν_1	2992.0	3111.1	3036.4	3010.0	3004.2	2989.1	-101.1	-106.9	-122.0
A'	ν_2	1473.6	1512.8	1487.5	1470.6	1469.7	1465.4	-42.2	-43.1	-47.4
A'	ν_3	1353.3	1379.3	1364.2	1348.6	1348.0	1346.9	-30.7	-31.3	-32.4
A'	ν_4	1067.8	1087.5	1063.4	1060.4	1060.0	1060.9	-27.1	-27.5	-26.6
A'	ν_5	759.9	781.2	773.9	769.1	763.4	768.9	-12.1	-17.8	-12.3
A'	ν_6	383.5	388.3	385.7	383.9	383.9	383.2	-4.4	-4.4	-5.1
A"	ν_7	3035.4	3190.1	3080.9	3049.9	3044.7	3043.4	-140.2	-145.4	-146.7
A"	ν_8	1236.8	1264.1	1241.5	1235.4	1234.9	1234.7	-28.7	-29.2	-29.4
A"	ν_9	1000.8	1022.4	1012.8	1005.3	1005.1	1004.8	-17.1	-17.3	-17.6
MAE				16.9	7.0	5.4	5.4			
			CC	DFT		CC/DFT			shift	
Symmetry	Mode	Exp.	Harm.	GVPT2	HYB	GHYB	HDHYB	HYB	GHYB	HDHYB
A'	ν_1	2992.0	3111.1	3015.7	3010.3	3009.2	2992.5	-100.8	-101.9	-118.6
A'	ν_2	1473.6	1512.8	1476.2	1473.0	1472.7	1468.3	-39.8	-40.1	-44.5
A'	ν_3	1353.3	1379.3	1352.6	1349.4	1350.0	1349.0	-29.9	-29.3	-30.3
A'	ν_4	1067.8	1087.5	1045.3	1059.7	1061.5	1061.5	-27.8	-26.0	-26.0
A'	ν_5	759.9	781.2	735.1	765.1	764.6	770.4	-16.1	-16.6	-10.8
A'	ν_6	383.5	388.3	377.7	383.8	384.2	383.5	-4.5	-4.1	-4.8
A"	ν_7	3035.4	3190.1	3053.2	3047.9	3047.0	3045.7	-142.2	-143.1	-144.4
A"	ν_8	1236.8	1264.1	1233.4	1235.6	1235.8	1235.6	-28.5	-28.3	-28.5
A"	ν_9	1000.8	1022.4	1000.7	1006.4	1007.2	1007.0	-16.0	-15.2	-15.4
MAE				11.3	6.2	5.8	4.9			

Table 2.1: Experimental and computed anharmonic frequencies (cm^{-1}) of CH_2ClF. GVPT2 refers to the Generalized VPT2 model,[24,25] the HYB, GHYB and HDHYB models are presented in Section 2.1. Mean Absolute Error (MAE) computed with respect to the experimental[73] frequencies. CC refers to the CCSD(T)/CBS+CV+aug method, and is the method chosen for the harmonic part in every case; CC', MP2 and DFT refer respectively to the CCSD(T)/AVTZ, the MP2/AVTZ, and the B2PLYP/AVTZ methods. See Ref.[73] for further details.

Integration limits (cm^{-1})	Main bands	Exp.[a]	CC/DFT HYB	CC/MP2 HYB
700-800	ν_5	78.6	86.61	86.38
950-1175	ν_9,ν_4	144	162.37	162.13
1175-1290	ν_8	2.40	2.43	2.42
1290-1410	ν_3	25.14	25.65	25.25
1430-1550	ν_2,$2\nu_5$	1.51	1.01	1.20
2050-2170	$2\nu_4$,$\nu_3+\nu_5$	2.94	2.91	3.02
2380-2530	$\nu_3+\nu_4$,$2\nu_8$	0.22	0.55	0.58
2780-2850	$\nu_2+\nu_3$	0.36	0.30	0.30
2850-3115	$2\nu_2$,ν_1,ν_7	26.3	25.68	24.43
3950-4200	$\nu_1+\nu_9$,$\nu_7+\nu_9$,$\nu_4+\nu_7$	0.9	0.96	0.99
4220-4300	$\nu_7+\nu_8$	0.17	0.13	0.17
4310-4420	$\nu_1+\nu_3$,$\nu_3+\nu_7$	0.32	0.39	0.35
4420-4570	$\nu_2+\nu_7$	1.26	1.31	1.34
5800-5990	$2\nu_1$	0.81	1.00	0.80
5990-6100	$2\nu_7$	0.25	0.65	0.48
	MAE	0.00	1.95	1.94

Table 2.2: Integrated cross sections (km/mol) of CH_2ClF. CCSD(T)/CBS (CC) harmonic intensities empirically combined with anharmonic shift at the B2PLYP/AVTZ (DFT) and the MP2/AVTZ (MP2) levels. Mean Absolute Error (MAE) computed with respect to the experimental data. [a] Ref.[73]

2.2.2 The hybrid DFT/DFT' model

The hybrid approaches have been used for the calculation of the anharmonic frequencies and intensities of the eight halo-substituted hydrocarbons,[70] and of the gas-phase glycine conformers[71,77] already encountered in Chapter 1. All the details about such computations can be found in the original works, while an interesting contribution to the present discussion is the estension of the GHYB approach from the CC/DFT level to the DFT/DFT' level, allowing to sensibly increase the dimension of the systems under study.

In fact, for all the 81 modes of the halo-hybrocarbons it was found[70] that the correction of the harmonic frequencies at the CC level are more important for the B3LYP/SNSD computations than for the B2PLYP/cc-pVTZ(-PP) ones, due to the already noticeable accuracy of the latter. This is graphically shown in Figure 2.1, where the signed errors for the B3LYP/SNSD and B2PLYP/cc-pVTZ(-PP) harmonic frequencies with respect to the CC ones, are plotted. The error distribution of the B2PLYP/cc-pVTZ(-PP) harmonic frequencies is much narrower than that for B3LYP/SNSD, the former being mostly within 10 cm^{-1}. This suggests that hybrid approaches in which the harmonic frequencies are computed at the B2PLYP/cc-pVTZ(-PP) level and the anharmonic shifts at the B3LYP/SNSD level – i.e. not requiring coupled-cluster computations –, can be used for large molecules containing halogens. The comparison between the MAE associated to the full DFT, DFT'/DFT and CC/DFT frequencies is shown in Table 2.3, for the

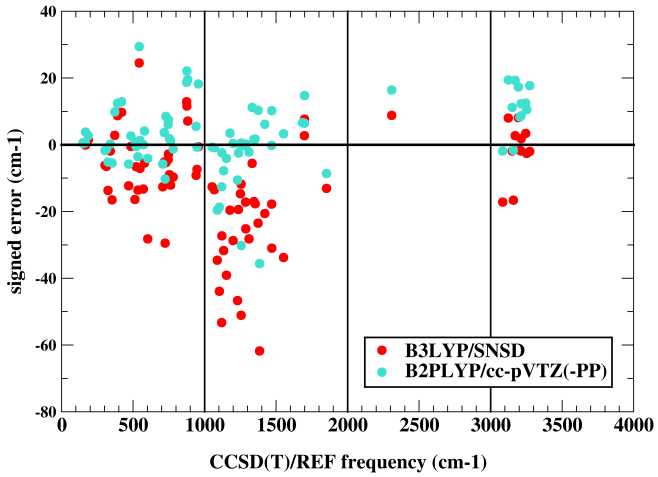


Figure 2.1: Signed errors (cm^{-1}) of the B3LYP/SNSD and B2PLYP/cc-pVTZ(-PP) harmonic frequencies with respect to the reference CCSD(T) ones.[70]

halo-hydrocarbons and for the glycine conformers discussed in Chapter 1. In the cases in which the B3LYP/DZP methods gives large errors, as is evident for the halo-methanes, the correction of the frequencies at the B2PLYP/TZP level can sensibly reduce the MAE. Such a correction can be useful also in order to reduce the maximum discrepancies on the single modes, as suggested by the maximum errors of the halo-methanes, halo-ethylenes, GLYIIn and GLYIIIp systems.

	MAE			MAX		
	DFT	DFT'/DFT	CC/DFT	DFT	DFT'/DFT	CC/DFT
Halo-methanes[a]	22.0	8.3	3.3	52.0	22.0	14.0
Halo-ethylenes[a]	9.3	9.4	4.7	41.0	32.0	24.0
GLYIp[b]	11.0	10.0	8.0	38.0	43.0	20.0
GLYIIn[b]	15.8	13.9	7.0	40.7	35.8	26.0
GLYIIIp[b]	12.0	13.0	4.0	34.0	28.0	20.0

Table 2.3: Maximum (MAX) and Mean Absolute (MAE) errors in cm^{-1} with respect to the experimental frequencies. DFT, DFT' and CC refer respectively to the B3LYP/DZP, B2PLYP/TZP and the CCSD(T)/TZP methods. See the original works for further details. [a] Ref.;[70] [b] Refs.[71,77]

Chapter 3

Reduced dimensionality VPT2 approach

In order to treat large molecules at the anharmonic level at a reasonable computational cost, it is possible to restrict the VPT2 treatment to a small part of the system, based on the spectroscopic observable of interest (for instance the most intense bands of an IR spectrum), by reducing the number of normal modes to be treated at the anharmonic level. Such an approach, which will be referred to as the Reduced Dimensionality VPT2 (RD-VPT2) in the following, provides a route to extend the anharmonic treatment to large molecules, and, contrary to the hybrid approach, has the strong advantage to employ a unique PES for all the computations. Nevertheless, the two methods can also be combined together for the treatment of complex macro-molecular systems. In this Chapter the RD-VPT2 approach for the calculation of anharmonic frequencies is first introduced, then, effective computational strategies for the selection of the normal modes to be included in the RD-VPT2 treatment are presented, and finally a practical use of the method is discussed. For further readings, the interested reader is referred to Refs.[50,51] from which the present Chapter is taken.

3.1 Theory

We refer to the formalism presented in Chapter 1 for the basic theory and implementation of the VPT2 approach. Let us first define a set of M_0 normal modes for which we want to compute the anharmonic frequencies. Such modes will be labelled as the *active modes* in the following, and for a non-linear[a] molecule composed by N atoms, $M_0 \leq 3N-6$. We will also refer the modes not included in the set as the *inactive modes*. The numerical differentiation of the Hessian matrix will be then performed only along the M_0 *active modes*, rather than the full set of $3N-6$ normal modes, so that $2M_0+1$ Hessian matrices, $\mathbf{H}^m$ ($m \in [1, 2M_0+1]$ is

[a] $M_0 \leq 3N-5$ for linear molecules.

a unique identification label), will be computed. Then, assuming that index i runs over the *active modes* ($i \in [1, M_0]$), and j and k are the indeces of the $\mathbf{H}^m$ matrices ($j, k \in [1, 3N-6]$), the cubic (k_{ijk}) and the semi-diagonal quartic (k_{iijk}) force constants will be evaluated. This can be practically performed in the GAUSSIAN code[133] by adding the SELECTANHARMONICMODES option to the FREQ keyword in the *route section* of the input, and then specifying the *active modes* after the *molecule specification section*.

Since in general $M_0 < 3N-6$, there will be a certain number of force constants which will be neglected with respect to the full dimensionality treatment, corresponding to the numerical differentiation of the Hessian matrix with respect to the *inactive modes*. Thus, when the RD-VPT2 is used for the computation of the anharmonic frequency of a given *active mode* i, the two contributions

$$\Delta_j = \frac{k_{jjj}k_{iij}}{4\omega_j} \quad \text{and} \quad \Delta_{jk} = \frac{k_{iik}k_{jjk}}{4\omega_k} \tag{3.1}$$

to the anharmonic matrix element χ_{ij} (Eq. 1.25) would be taken into account only if j for Δ_j and both j and k for Δ_{jk} are *active modes*, otherwise they would be neglected, k_{jjj} and k_{jjk} being unknown.[50] Since both terms are non-resonant, the Fermi resonances, if any, remain unchanged.

If, for a given set of M_0 *active modes*, the missing Δ_j and Δ_{jk} are small enough, the anharmonic frequencies computed with the RD-VPT2 approach will be very similar to the full VPT2 ones. If the Δ_j and Δ_{jk} of an *inactive mode* are large, the RD-VPT2 will be affected by large errors with respect to the standard VPT2. In this case, such a mode should be inserted in the set of the *active modes*, in order to converge to the VPT2 anharmonic frequencies, taking into account the corresponding Δ_j and Δ_{jk}. For this reason, in principle, the number of *active modes* which need to be included in the RD-VPT2 calculations (M_1) can be larger than the set of anharmonic frequencies which are under study (M_0), so that $3N-6 \geq M_1 \geq M_0$. As a consequence, the choice of the optimal set of *active modes* is one of the most crucial steps in the RD-VPT2 appproach.

3.1.1 Selection of the *active modes*

The RD-VPT2 approach has been developed for large systems, for which, in principle, the full VPT2 calculations are not practical. Thus, one of the most important point is *how could we predict that the terms Δ_j and Δ_{kj} in Eq. 3.1 are small enough to be neglected, the magnitude of many force constants being unknown a priori?*

One possible approach – the one which has been employed in the following – is to proceed iteratively by first choosing a set of M_0 *active modes* and computing the corresponding cubic force constants. Indeed, it is possible to estimate the magnitude of Δ_j and Δ_{kj} by evaluating the ratio k_{iij}/ω_j, since k_{iij} is known at this point. If such a ratio is small, it is reasonable to assume that both the Δ_j and the Δ_{kj} will be small enough to be neglected. A larger value of k_{iij} reflects a significant coupling

between modes i and j, meaning that mode j should be also included in the set of *active modes*. If the ratio k_{iij}/ω_j is not small enough to safely exclude mode j and not large enough to suggest an automatic inclusion in the set of *active modes*, a numerical differentiation of the Hessian matrix along mode j could be performed, in order to explicitely compute k_{jjj} and k_{jjk} force constants. This is still much less computationally demanding than the full VPT2 approach.

A general procedure for the selection of the normal modes to be included in the RD-VPT2 treatment can be summarized in the following steps:

1. Define a set of M_0 normal modes which are of interest for the specific problem under study. For example, it can be defined by selecting a portion of an Infrared or Raman spectrum, or for the analysis of specific vibrations of a molecule related to a fragment or to a subsystem of the whole system. Chemical intuition can be helpful in this step, in order to guess the possible couplings between the modes, e.g. if, say, a CH_2 asymmetric stretching is included in the *active modes*, it is very likely to happen that a strong coupling occurs with the corresponding symmetric mode.

2. Compute the anharmonic frequencies obtained with the M_0 *active modes*, using the RD-VPT2 approach.

3. The rectangular matrix ($M_0 \times 3N-6$) of k_{iij} coupling constants is then available. Analysing the values of the off-diagonal k_{iij} terms, determine which are the largest constants which could give significant contributions to the anharmonic frequencies of the M_0 selected modes.

4. Define a new set of *active modes* (M_1) adding the ones which have been found to be strongly coupled with the modes in M_0 ($3N-6 \geq M_1 \geq M_0$), compute the missing force constants by differentiation of the Hessian matrix along the new modes, and evaluate again the anharmonic frequencies.

5. Iteratively repeat the last two steps until the M_0 frequencies of interest show very small refinements from a smaller to a larger ensemble of *active modes*.

It is worth noting that full dimensional VPT2 computations are never required by the procedure of selection of the normal modes. This is very important when the RD-VPT2 treatment is applied to large systems for which they are not computationally conceivable. On the other side, an important drawback of this approach is that the computational cost is strongly dependent on the specific chemical system under study. In fact, the more the normal modes are coupled with each other, the larger is the number of modes which need to be included in the computations. At most, if all the modes are strongly coupled, the approach requires all the normal modes to be active, resulting in the RD-VPT2 to need the same computational time and memory as the standard VPT2 calculations. Conversely, for systems in which the couplings are small, the RD-VPT2 approach allows to sensibly decrease the computational

demand of the calculations. Typical examples of such systems are those composed of organic molecules strongly – covalently – adsorbed on inorganic crystal sufaces, for which the vibrations of the molecule are usually weakly coupled with the surface phonons.

3.1.2 Numerical tests

In the following the most stable conformer of the gas-phase glycine molecule (GLYIp, see Figure 1.5 and Section 1.2.3) will be used as a test case in order to illustrate the most important aspects of the RD-VPT2 approach. We will discuss the computation of the anharmonic frequencies of five stretching modes (i.e., modes 20-24) of GLYIp, although the discussion is very general and the same arguments can be in principle extended to every normal mode of every molecular system.

In Table 3.1, the anharmonic frequencies for modes 20-24 calculated with the standard VPT2 approach and with different RD-VPT2 schemes are reported. The standard VPT2 approach can be viewed as a RD-VPT2 approach in which all the normal modes are treated at the anharmonic level, so we will refer to such calculations as the "All-Modes" approach (AM). In the 1M approach, we computed the frequencies of modes 20-24 treating only one mode – the mode of interest – at the anharmonic level. This is very useful when the mode under study is intrinsically anharmonic, and the contributions from the couplings with the other modes are small. This is also the cheapest way to compute anharmonic frequencies within the RD-VPT2 scheme. The couplings with the other modes have then been included in the five (5M) and six (6M) modes approaches. In the 5M approach the anharmonic frequencies of the five stretching modes were calculated including only the five modes under study (i.e., modes 20-24). In the 6M approach mode 10 was also added. All the calculations were performed at the B3LYP/aug-N07D level, and the experimental frequencies are also reported for comparison. All the diagonal k_{iii} and off-diagonal k_{iij} cubic force constants are known from the AM calculations, so and they are graphically represented in Figure 3.1, in order to better analyse the results of the RD-VPT2 approaches. When we perform the 1M calculation for mode 24 we numerically differentiate the Hessian matrix along the one-dimensional PES of the mode, so the first row in Figure 3.1 is known. Since also the harmonic frequencies are known, we can estimate all the ratios $k_{24,24,j}/\omega_j$. This mode is highly anharmonic, so $k_{24,24,24}$ is very large (2570 cm^{-1}), while the sum of all the off-diagonal terms k_{iij} related to the coupling with the other modes is only 289 cm^{-1}. The biggest off-diagonal cubic force constant is $k_{24,24,14}$ and is only 96 cm^{-1}. Thus, the ratio $k_{24,24,j}/\omega_j$ will be lower than 1 for every j, and it is reasonable to assume that all Δ_j and Δ_{kj} are small enough to be neglected. This can be *a posteriori* confirmed by the 2M calculations including mode 14 with 24, which improve the 1M frequency of only 3 cm^{-1}, and by the fact that the 1M frequency differs by just 7 cm^{-1} from the AM one.

1M computations for mode 23 and 21 provide the entire set of the $k_{23,23,j}$ and

$k_{21,21,j}$ coupling constants (the second and fourth rows in Figure 3.1). The intrinsic anharmonicity for these modes is very small ($k_{23,23,23}$ and $k_{21,21,21}$ are both close to zero), while the coupling effects are very important (the sums of the off-diagonal terms, k_{iij}, are 2255 cm^{-1} and 1647 cm^{-1}, respectively). Thus, it is evident that the 1M approach may not be sufficient for the computation of the anharmonic frequencies of these two modes, and a certain number of couplings should be included. The largest off-diagonal coupling constants are $k_{23,23,22}$ (1645 cm^{-1}) for mode 23 and $k_{21,21,20}$ (1405 cm^{-1}) for mode 21. This type of coupling was expected for these modes, since it can be physically viewed as the coupling between the symmetric and asymmetric stretching modes of the -NH_2 (modes 23 and 22) and of the -CH_2 (modes 21 and 20) moieities of the glycine molecule. The 5M approach includes all these couplings, and in fact strongly changes the frequencies of modes 23 and 21 with respect to the 1M ones.

Furthermore, regarding mode 23, Figure 3.1 shows that another important coupling is connected to $k_{23,23,10}$, and the inclusion of mode 10 in the VPT2 treatment further changes the frequency by 10 cm^{-1} (see the 6M result in Table 3.1). Other relevant coupling constants are $k_{23,23,8}$ and $k_{23,23,18}$, but no significant changes were found when such modes were included in the treatment. Thus, the anharmonic shift can be mainly attributed to the terms depending on $k_{23,23,22}$ and $k_{23,23,10}$, and the 6M treatment gives a very good estimation of the AM value. This can be *a posteriori* confirmed by the fact that the 6M frequency differ by only 4 cm^{-1} from the AM value and by only 1 cm^{-1} from the experimental value. On the other side, for mode 21, although $k_{21,21,17}$ is a large constant, test calculations showed that it actually has a negligible effect, and it can be *a posteriori* verified that the 5M scheme is in agreement within 2 cm^{-1} with the AM frequency value.

Mode i:	24	23	22	21	20	
Assignment	$\nu(OH)$	$\nu(NH_2)a$	$\nu(NH_2)s$	$\nu(CH_2)a$	$\nu(CH_2)s$	MAE
Exp.	3585[a]	3410[b]	3357[c]	-	2939[c]	-
AM[d]	3568	3407	3387	2929	2938	10
6M[e]	(-7)	(+4)	(-33)	(+1)	(-25)	11
5M[f]	(-7)	(-14)	(-45)	(+2)	(-25)	16
1M[g]	(-7)	(-359)	(-42)	(-70)	(-22)	84
k_{iii}	2570	$\sim$0	1617	$\sim$0	1345	
$\sum \lvert k_{iij} \rvert$	289	2255	302	1647	134	
$\lvert k_{iij} \rvert_{\max}$	96	1645	114	1405	25	
j_{max}	14	22	10	20	16	

Table 3.1: Anharmonic vibrational frequencies and force constants (in cm^{-1}) of five stretching modes of the GLYIp conformer of the gas-phase glycine molecule. Experimental frequencies are also shown for comparison. "a" and "s" stand for "asymmetric" and "symmetric" stretching modes, respectively. In parenthesis the differences with respect to the AM frequencies have been reported. [a] Ref.;[163] [b] Ref.;[5] [c] Ref.;[4] [d] All-modes included; [e] Modes 20-24 and 10 included; [f] Modes 20-24 included; [g] Only the mode under study was included (e.g., only mode 24 included in the calculation of the frequency of mode 24).

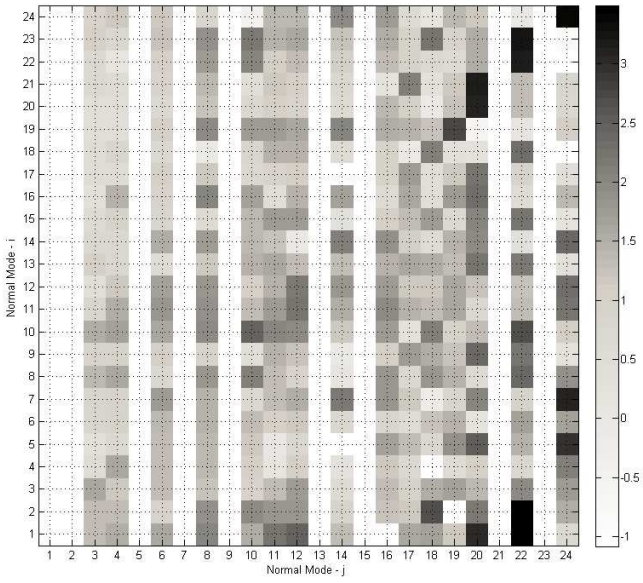


Figure 3.1: Graphical representation of the absolute value of the cubic force constants ($|k_{iij}|$) for the GLYIp conformer of the gas-phase glycine molecule. A shade of grey is assigned depending on the value of $\log_{10}(|k_{iij}|)$, from white for the values lower than 0.1 to black for values above 3.

3.2 Glycine adsorbed on the (100) Silicon surface.

The RD-VPT2 approach together with the selection procedure of the normal modes has been applied for the analysis of the High Resolution Electron Energy Loss (HREEL) spectrum of the glycine molecule adsorbed on the (100) surface of a Silicon crystal, which was measured by Lopez *et al.*[175] in ultra-vacuum conditions at different temperatures. Although neither Infrared nor Raman intensities could be compared with the HREEL spectral lineshapes, the position of the peaks can be compared with the anharmonic frequencies. As we will see in the following, this comparison can be very useful either to provide insights in the assignment of the spectrum, or to interpret the experimental data in terms of an atomistic picture. First, the adsorption process of the glycine molecule on the Si(100) surface was studied; then, the vibrational frequencies of the adsorbate were compared with the peak positions of the HREEL spectrum. The B3LYP/aug-N07D[144,151–153] level of theory has been used throughout this work.

3.2.1 The chemical system

The (100) surface of the Silicon crystals is by far the most studied one.[182–187] It is known that it undergoes a reconstruction from the bulk crystal structure which leads to the formation of rows of buckled dimers. In such a buckled structure the two silicon atoms are polarised, the lower one showing an electrophilic character, the upper one a nucleophilic character.[166,188] A strong σ bond and a π interaction between the upper and lower silicon atoms determine the electronic structure of the dimers and these features are key factors for the surface reactivity.[188,189] For further details about Si(100) surface properties we refer the interested reader to review works.[189,190] The adsorption of glycine on such a surface is likely to be determined by the local interaction between the molecule and the dimers, so we developed three cluster models of the Si(100) surface, represented in Figure 3.2, taking into account one or more dimers. The minimal cluster has stoichiometry Si_9H_{12} and it exposes one dimer on the surface, while $Si_{15}H_{16}$ has two dimers on the surface, and both of them have been already used in many other previous studies.[20,176,186,187,191–193] The $Si_{60}H_{44}$ model exposes four dimers, taking into account an extended portion of the surface, and to the best of our knowledge it has never been used before this study.

Regarding the glycine molecule, as discussed in Section 1.2.3, the most stable glycine conformer (GLYIp in Figure 1.5) has a relative abundance of more than 70% in the gas-phase mixture,[4,5] so that it is reasonable that this conformer is also the one which most probably binds the silicon surface under the experimental conditions reported by Lopez *et al.*[175] In previous theoretical studies of the adsorption process of glycine on Si(100) it has been shown[176,177,194] that while both bindings via the $-NH_2$ and the -COOH moieties of glycine lead to energetically stable structures, the -COOH binding path should be more probable because of its much lower activation energy. Indeed, this seems to be confirmed by the HREEL spectrum,[175] where the

features of the OH stretching are missing. For this reason, we simulated the adsorption process via the -COOH binding, computing some relevant thermodynamical and energetical quantities. When GLYIp approaches the silicon dimer from infinite distance, a local minimum is formed with an energy gain of about 20 kcal/mol with respect to the two isolated subsystems. The local minimum, which will be referred to as Ads(COOH) in the following, is shown in Figure 3.3, while the adsorption energies are reported in Table 3.2. In Ads(COOH) the Si2-O2 distance is relatively

		Ads(COOH)	TS	Ads(COO+H)	Activation Energy	Reaction Energy
GLYIp@Si_9H_{12}	E_{ads}	-20.24	-20.15	-54.44	0.09	-34.20
	E_{ads}^{BSSE}	-18.91	-18.77	-53.86	0.14	-34.95
	$E_{ads}^{BSSE,ZPVE}$	-19.51	-21.09	-54.89	-1.57	-35.37
GLYIp@$Si_{15}H_{16}$	E_{ads}	-22.60	-21.91	-53.51	0.69	-30.91
	E_{ads}^{BSSE}	-21.16	-20.38	-52.05	0.78	-30.89
	$E_{ads}^{BSSE,ZPVE}$	-20.93	-22.77	-52.89	-1.84	-31.96
GLYIp@$Si_{60}H_{44}$	E_{ads}	-23.35	-	-53.19	-	-29.84

Table 3.2: Energetics (kcal/mol) of the adsorption process computed for different clusters. The Activation Energy is the energy difference between E_{ads} of TS and Ads(COOH); the Reaction Energy is the energy difference between E_{ads} of Ads(COO+H) and Ads(COOH). E_{ads} is the adsorption energy computed as difference between the energy of the interacting fragments and the sum of the energy of the isolated fragments. BSSE and ZPVE corrections have been also included.

short (1.894 Å for the Si_9H_{12} cluster), and the Mulliken Atomic Charges (MAC) of O2 and Si2 pass respectively from -0.686 and $+0.369$, in GLYIp to -1.314 and $+1.740$, in Ads(COOH), suggesting that the π electronic density of the silicon dimer has been partially moved to the molecule. Generally speaking, the adsorption energy of Ads(COOH) is largely due to the Si2-O2 and Si1-H1 interactions. In the subsequent step of the adsorption process, the OH dissociation occurs, and the H1 proton is transferred to the Si1 atom, leading to the dissociated product shown in Figure 3.4, which will be referred to as Ads(COO+H). The Transition State (TS) of the Ads(COOH)→Ads(COO+H) process has been calculated, and a large imaginary frequency has been found for the OH stretching mode (see Figure 3.3). From Table 3.2, it can be observed that the OH dissociation is a spontaneous reaction, with an Activation Energy lower than 1 kcal/mol, and the BSSE and ZPVE corrections do not qualitatively change this result. Then, the whole energy profile of the OH dissociation process has been evaluated from *Hessian-based Intrinsic Reaction Coordinate*[195,196] (IRC) calculations on the Si_9H_{12} cluster, and reported in Figure 3.5. It has been observed that such a process can be divided into two separate steps: the first one (values from 0 to 8 of the reaction coordinate) involves the proton transfer reaction, whereas the second one (from 8 to 24) corresponds to a rotation of the molecule about the Si2-O2 axis (Figure 3.4). The overall reaction energy of the OH dissociation is 34.2 kcal/mol, about 31.4 kcal/mol related to the proton transfer and

2.8 kcal/mol to the subsequent rotation. The adsorption energies "E_{ads}" reported in Table 3.2 relative to the system GLYIp@Si_9H_{12} have been plotted in Figure 3.6 to provide a graphical representation of the energy profile of the entire adsorption process, while further details about the reaction mechanism can be found in the original work.[51]

3.2.2 Vibrational analysis

The interpretation of the vibrational spectra is often a non-trivial task, which in the case of the experimental HREEL spectrum reported by Lopez *et al.*[175] is additionally complicated by the presence of many convoluted peaks contributing to the band broadening. For this reason it is difficult to exactly distinguish the transitions due to the glycine molecule from the transitions due to the surface phonons, and quantum mechanical methods provide useful tools to unravel the complexity of the spectrum. Indeed, in the original work the assignment of the transitions has been assisted by DFT calculations, to obtain the frequencies at the harmonic level, and employing a cluster with stoichiometry Si_8H_{10} as a model of the surface. In the present work, we have computed the frequencies using the RD-VPT2 approach in conjunction with the B3LYP/aug-N07D force field, on the Ads(COO+H) geometry (Figure 3.4), using the GLYIp@Si_9H_{12} and the GLYIp@$Si_{15}H_{16}$ model systems, and treating the Fermi resonances with the GVPT2 model.[24,25] Harmonic computations were first performed in order to assign all the normal modes, and the modes of the adsorbed glycine molecule were selected as a first set of *active modes*. Then, the force constants were analysed and the modes significantly coupled with those *active modes* were added to the treatment, while the modes which did not give any significant contribution to the spectrum were removed, leading to a final set of 19 *active modes* for GLYIp@Si_9H_{12} and 26 modes for GLYIp@$Si_{15}H_{16}$. In particular, in both GLYIp@Si_9H_{12} and GLYIp@$Si_{15}H_{16}$ model systems, the largest coupling constant between the molecule and the cluster modes was related to the SiH stretching, with values of 142 cm^{-1} and 206 cm^{-1}, respectively. Furthermore, similarly to the results of the isolated glycine reported in Section 3.1.2, the NH_2 asymmetric stretching is characterized by a low intrinsic anharmonicity and a strong coupling with the symmetric stretching, while the latter has a strong intrinsic anharmonicity. A similar behaviour is also found for the CH_2 stretching modes. Magnitudes lower than 120 cm^{-1} have been found for the k_{iij} related to all the other modes, suggesting that such couplings could be safely neglected. More details can be found in the original works.[50,51]

In Table 3.3 and 3.4 the vibrational frequencies of the glycine and of deuterated glycine, respectively, adsorbed on the Si(100) are reported, as experimentally estimated by Lopez *et al.*,[175] as computed by Shemesh *et al.*[20] with the VSCF and cc-VSCF methodologies[17,19,21,32,33] using a semiempirical (PM3) potential energy surface, and as computed in this work with the RD-VPT2 method. Because of the broad-band structure in the low energy range of the spectrum, an unambiguous

comparison between the experimental, the RD-VPT2 and the (cc-)VSCF frequencies can only be ensured in the high-energy range of the spectrum, so that the Mean Absolute Error (MAE) with respect to experimental data has been calculated only for the highest-frequency modes.[b] The differences between the frequencies computed with the two model clusters for the non-deuterated case are lower than 13 cm^{-1} for all the modes, with the only exception of mode 4 (CH_2 symmetric stretching), where a discrepancy of 39 cm^{-1} is observed. For the deuterated glycine, the maximum difference between the two model clusters is related to the SiD stretching mode, and is lower than 12 cm^{-1}. Thus, consistent results between the two model clusters have been obtained.

In Table 3.3, the MAE of the RD-VPT2 frequencies is 34 cm^{-1} for GLYIp@Si_9H_{12} and 38 cm^{-1} for GLYIp@$Si_{15}H_{16}$, but in both cases an error larger than 100 cm^{-1} is found for the frequency of mode 3 (CH_2 asymmetric stretching), and by excluding such a mode, the two MAEs decrease to 22 cm^{-1} and 26 cm^{-1}, respectively. In particular, the frequency of mode 3 is about 2931 cm^{-1} at the RD-VPT2 level, for both the cluster models employed, while the experimental estimation is 3040 cm^{-1}. VSCF and cc-VSCF calculations are in closer agreemeent with the experimental estimation for the smaller cluster model, while they are in closer agreement with the RD-VPT2 prediction for the larger one. As a consequence, some more data should be evaluated in order to correcly assign such a band. Looking at the experimental HREEL spectra reported in the original work,[175] the band at 3040 cm^{-1} is still observed in the HREEL spectrum at 200° C, where evidences of desorption of the glycine from the surface are present, as well as at 400° C and 500° C where the glycine is considered completely desorbed. Additionally, the HREEL spectrum[8] of a system very similar to the one under analysis, [c] shows the feature related to the CH_2 asymmetric stretching of the adsorbed glycine at 2960 cm^{-1}, and it can not be excluded that a non-negligible presence of the zwitterionic form of the amino acid in the experimental sample contributes to the broadening of the peak. Also the HREEL spectrum[197] of ethylene adsorbed onto the Si(100)-(2×1) surface shows the CH_2 stretching band at 2950 cm^{-1}. Furthermore, the RD-VPT2 frequencies of the deuterated glycine reported in Table 3.4 show an overall MAE of 27 cm^{-1} for Si_9H_{12} and 29 cm^{-1} for $Si_{15}H_{16}$, and it is noteworthy that the asymmetric stretching frequency of CD_2 matches very well the experimental value, with discrepancies of only 17 cm^{-1} for Si_9H_{12} and 11 cm^{-1} for $Si_{15}H_{16}$. From these arguments, we suggest that the asymmetric stretching frequency of the CH_2 group of the adsorbed glycine should be more similar to the value of 2960 cm^{-1} reported by Lafosse *et al.*[8] than to the value of 3040 cm^{-1} reported by Lopez *et al.*[175] The broad feature at 3040 cm^{-1} in the spectrum of Lopez *et al.*[175] could be rather associated to the presence of zwitterionic forms of the glycine and/or to other species which are still present at

[b] In Table 3.3 seven modes have been considered, while six modes have been included in the MAE of Table 3.4 for the deuterated glycine.

[c] A glycine film synthetised onto a diamond layer deposited on a p-doped silicon substrate.

500° C.

Comparing the frequencies of the adsorbed glycine (Table 3.3) with the gas-phase GLYIp ones (Table 1.5), it can be noted that the CH_2 and NH_2 stretchings and the NH_2 bending modes differ by less than 15 cm^{-1}, suggesting that the interactions between the glycine molecule and the "back dimers" (see Figure 3.4) play a negligible role and that the GLYIp@Si_9H_{12} and GLYIp@$Si_{15}H_{16}$ are large enough to correcly represent the binding site. The frequency of the C=O stretching mode is reduced by 70 cm^{-1} when passing from the gas-phase GLYIp to GLYIp@Si_9H_{12} and GLYIp@$Si_{15}H_{16}$, in qualitative agreement with the experimental red-shift of about 30 cm^{-1} found by comparing the frequencies of the HREEL[175] and the IR[4,5] spectra. Such a red-shift is consistent with the geometrical changes found along the absorption path, and it could be attributed to the flow of charge from the cluster to the molecule occurring in the adsorption process.[51]

In Figure 3.7 and Figure 3.8, a graphical representation of the computed peak positions on the experimental HREEL spectrum is given.

		GLYIp@Si_9H_{12}			GLYIp@$Si_{15}H_{16}$			
Mode	Assignment	RD-VPT2 (19 modes)	VSCF[a]	CC-VSCF[a]	RD-VPT2 (26 modes)	VSCF[a]	CC-VSCF[a]	exp.[b]
1	NH_2 (A) str.	3421	3548	3548	3419	3391	3374	3460
2	NH_2 (S) str.	3381	3456	3457	3379	3305	3292	3410
3	CH_2 (A) str.	2931	3034	3033	2932	2927	2915	3040
4	CH_2 (S) str.	2940	2980	2980	2901	2865	2821	2940
5	Si-H str.	2096	1993	1992	2110	1992	1990	2115
6	C=O str.	1713	1722	1722	1712	1668	1660	1750
7	NH_2 bend.	1625	1669	1669	1619	1668	1666	1620
8	CH_2 GLYIp bend.	1424	-	-	1405	-	-	-
9	Twist 1	1341	1386	1385	1340	1370	1365	1380
10	CH_2 z bend.	1308	-	-	1313	-	-	-
11	C-O + CC str.	1219	1178	1177	1220	1226	1220	1260
12	Twist 2	1154	1194	1193	1157	1180	1177	1170
13	CN str.	1095	-	-	1101	-	-	-
14	CC str.	874	990	989	879	950	947	925
15	Twist 3	896	-	-	889	-	-	-
16	OCO bend.	788	-	-	780	-	-	-
17	Si-O str.	755	737	736	753	785	783	780
18	-	566	660	660	570	668	667	630
19	-	483	437	437	481	-	-	475
MAE		34	54	55	38	88	101	-

Table 3.3: Experimental and computed frequencies of glycine@Si(100). The Mean Absolute Error (MAE) has been calculated by considering only modes 1-7. RD-VPT2 calculations have been performed using the GVPT2 scheme for the treatment of the Fermi resonances. In the assignment column, "A" and "S" refer to "asymmetric" and "symmetric", respectively, "ip" stands for "in plane", "z" indicates the direction perpendicular to the (100) suface plane. [a] Shemesh *et al.*;[20] [b] Lopez *et al.*[175]

		D-GLYIp@Si_9H_{12}			D-GLYIp@$Si_{15}H_{16}$	
Mode	Assignment	RD-VPT2 (17 modes)	VSCF[a]	CC-VSCF[a]	RD-VPT2 (24 modes)	exp.[b]
1	ND_2 (A) str.	2555	2548	2532	2553	2584
2	ND_2 (S) str.	2452	2419	2400	2448	2520
3	CD_2 (A) str.	2217	2187	2174	2211	2200
4	CD_2 (S) str.	2141	2118	2094	2143	2112
5	C=O str.	1723	1651	1650	1721	1736
6	Si-D str.	1524	1461	1460	1535	1520
7	ND_2 bend.	1234	-	-	1232	-
8	C-O str.	1217	1212	1208	1212	1216
9	CN str.	1095	1122	1119	1096	1120
10	CH_2 z bend.	1056	-	-	1051	-
11	Twist 1	1036	-	-	1032	-
12	CH_2 ip bend.	953	-	-	954	-
13	CD_2 twist	930	960	958	933	936
14	OCO bend.	857	872	870	855	880
15	Si-O str.	755	740	736	750	736
16	Twist 2	722	-	-	713	-
17	ND_2 oop bend.	677	678	678	678	615
18	-	-	517	516	487	472
MAE		27	50	60	29	-

Table 3.4: Experimental and computed frequencies of deuterated glycine@Si(100). The Mean Absolute Error (MAE) has been calculated including only modes 1-6. RD-VPT2 calculations have been performed using the GVPT2 scheme for the treatment of the Fermi resonances. In the assignment column, "A" and "S" refer to "asymmetric" and "symmetric", respectively, "ip" and "oop" stand for "in plane" and "out of plane", respectively, "z" indicates the direction perpendicular to the (100) suface plane. [a] Shemesh *et al.*;[20] [b] Lopez *et al.*[175]

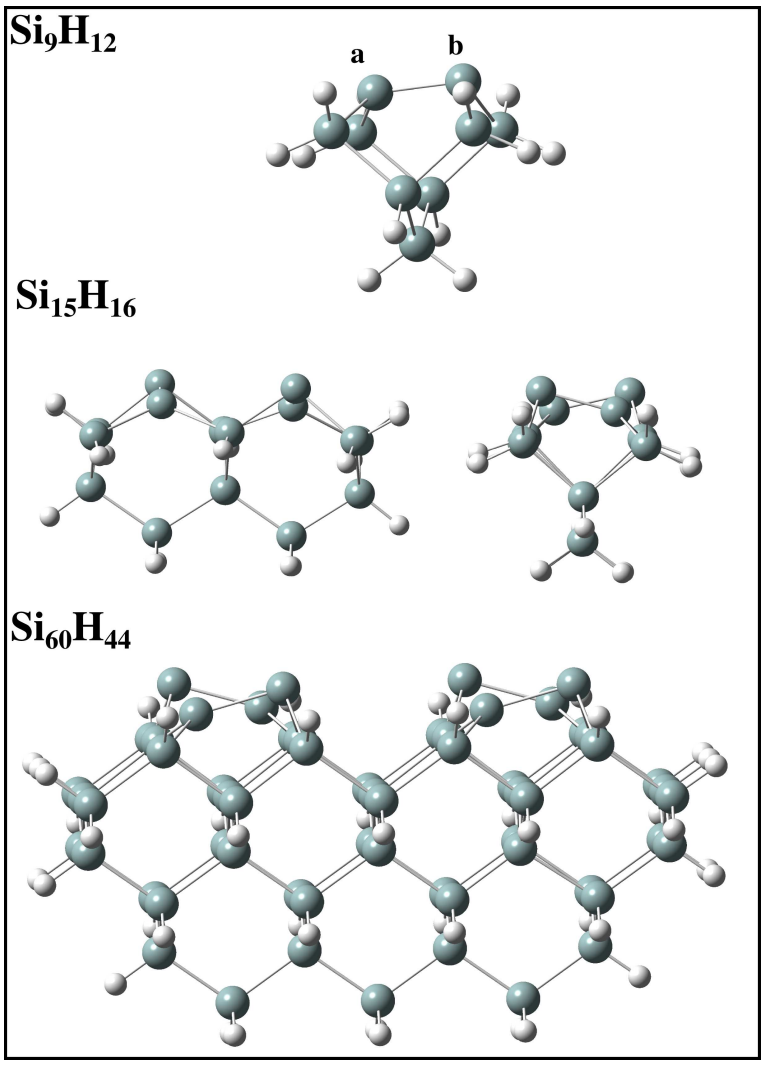


Figure 3.2: Clusters employed throughout this work. "a" and "b" indicate the atoms belonging to the dimer. The $Si_{15}H_{16}$ is shown from two different perspectives in order to evidence the buckling of the two dimers.

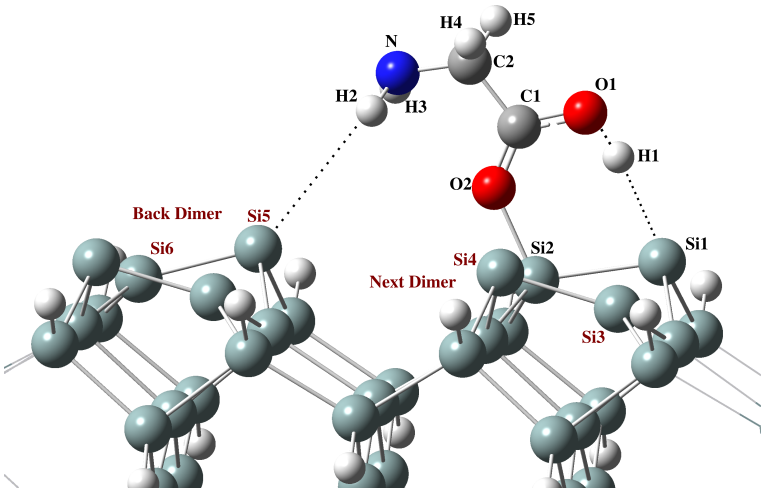


Figure 3.3: Ads(COOH) binding site on the $Si_{60}H_{44}$ cluster. This picture can be taken as reference also for the geometry of the Transition State of the OH dissociation process, since the two structures are very similar.

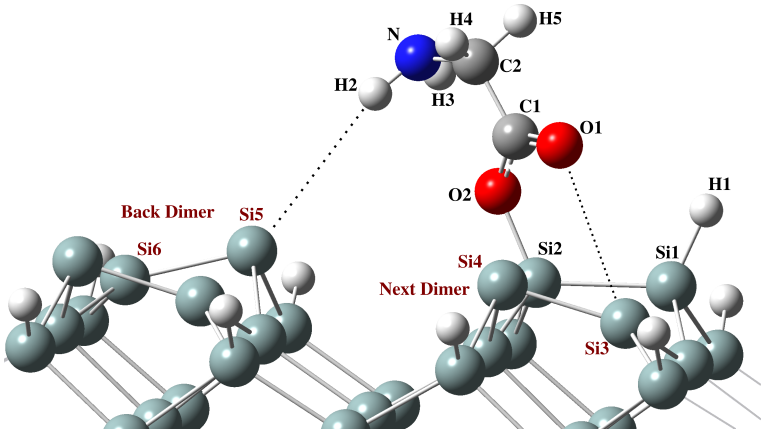
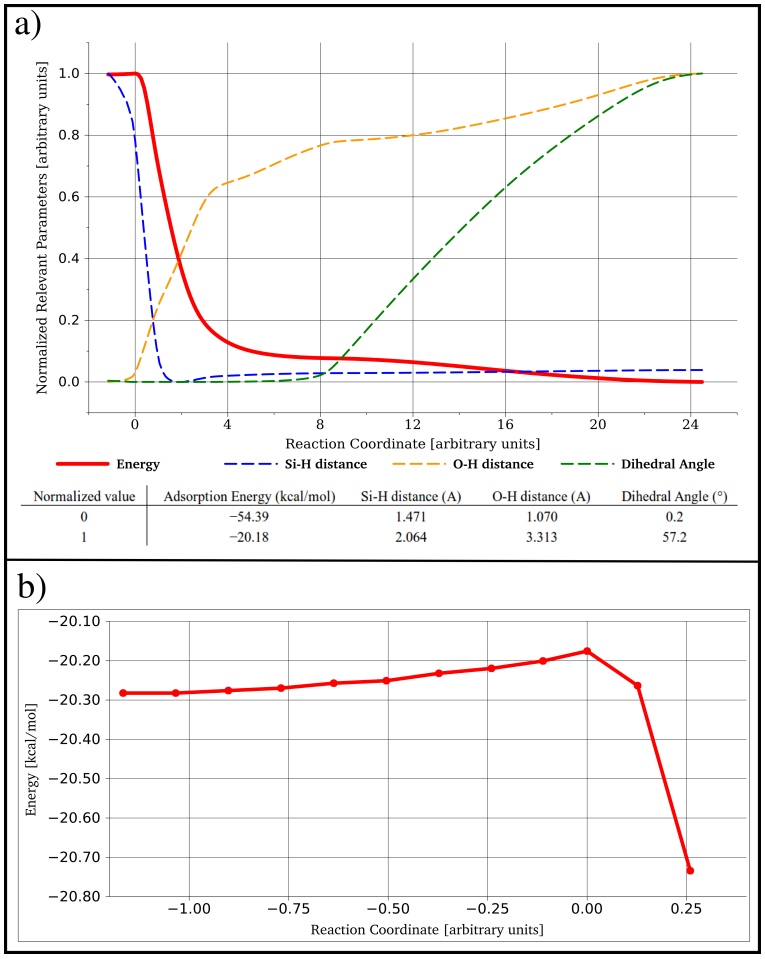


Figure 3.4: Ads(COO+H) binding site on the $Si_{60}H_{44}$ cluster.



Normalized value	Adsorption Energy (kcal/mol)	Si-H distance (A)	O-H distance (A)	Dihedral Angle (°)
0	−54.39	1.471	1.070	0.2
1	−20.18	2.064	3.313	57.2

Figure 3.5: Variation of the energy and of some selected geometrical parameters along the OH dissociation reaction path, resulting from *Hessian-based IRC* computations for GLYIp@Si_9H_{12} model system. In panel a) the energy, the bond-lengths and the angle have been normalized to unity. In panel b) the energy profile (in kcal/mol) in the zone of the Transition State is shown.

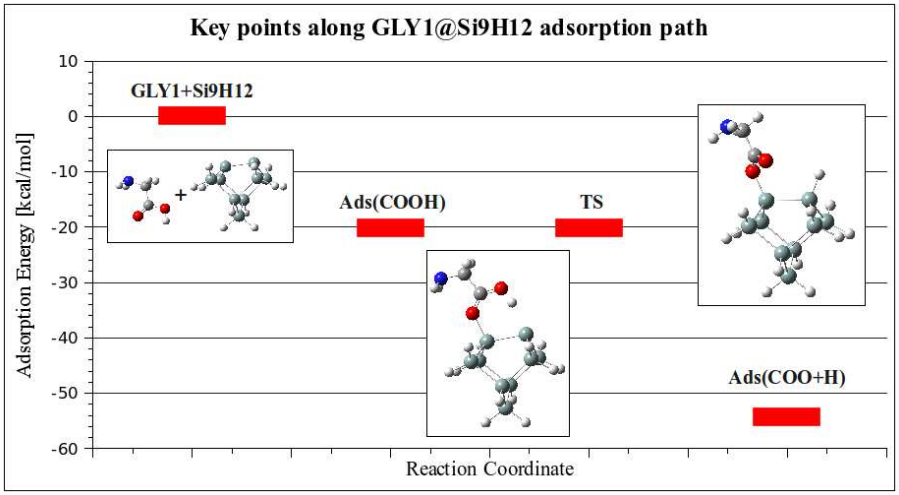


Figure 3.6: Geometries and energies of the stationary points of the adsorption path.

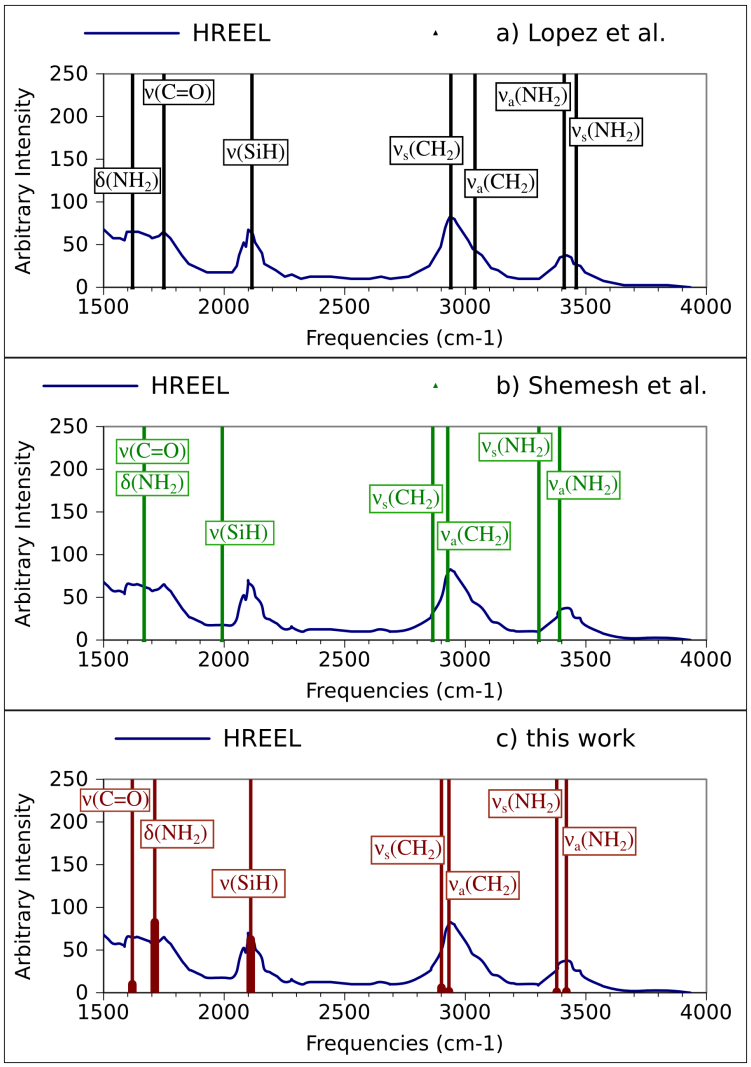


Figure 3.7: Experimental[175] (a), VSCF[20] (b) and RD-VPT2 (c) frequencies (vertical lines) superimposed to the HREEL spectra[175] in the 1500-4000 cm^{-1} frequency range.

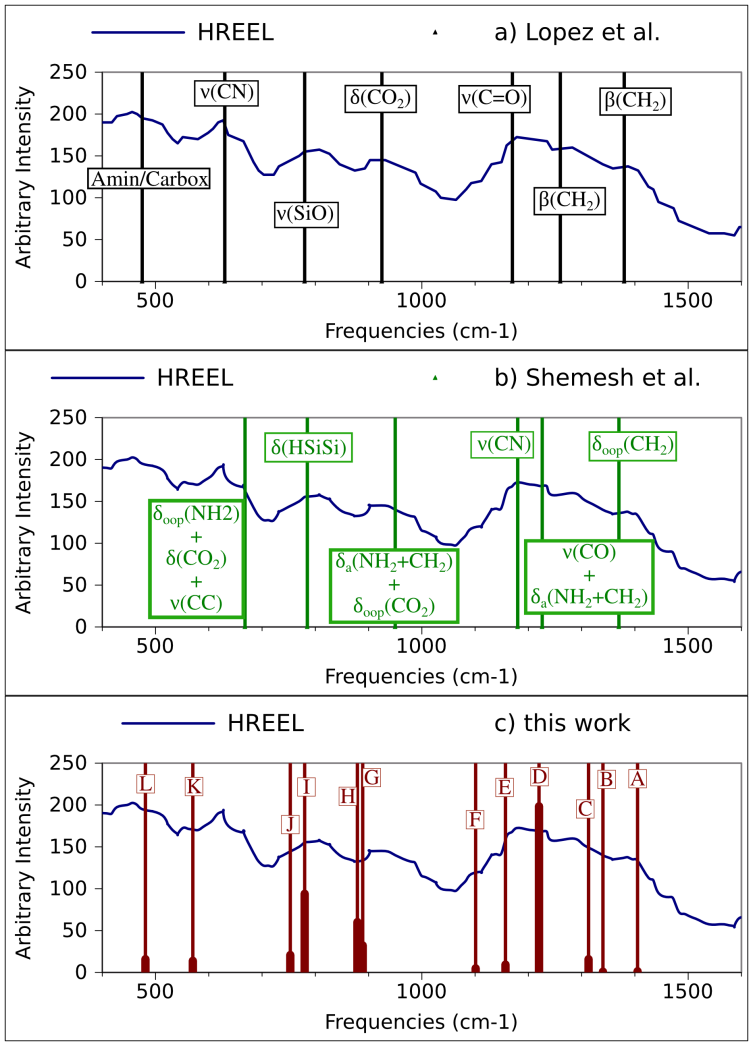
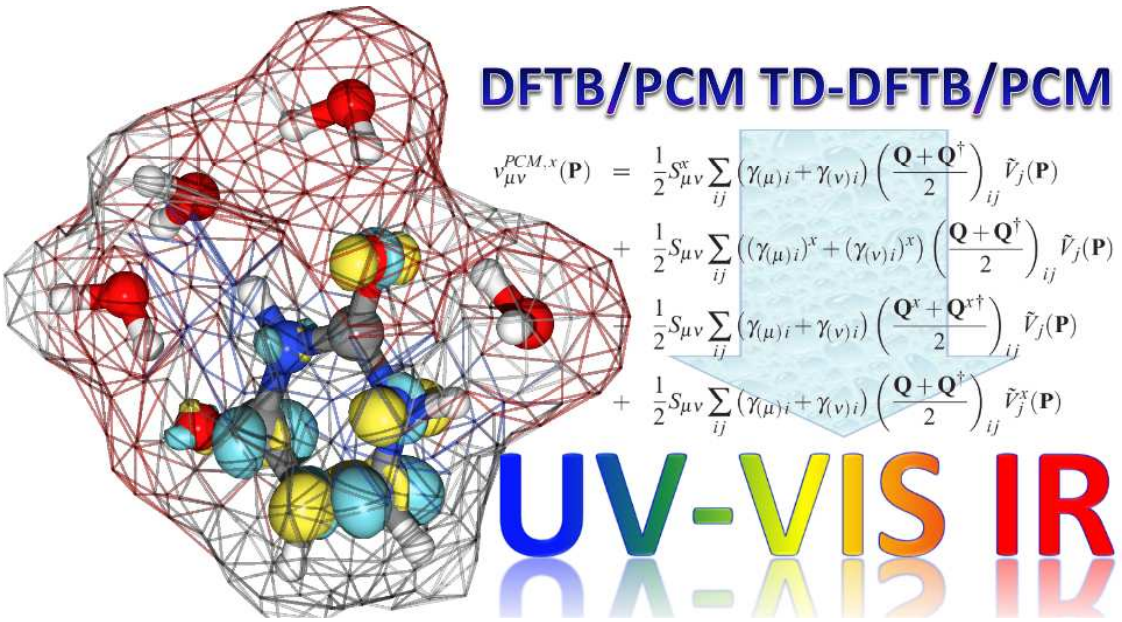


Figure 3.8: Experimental[175] (a), VSCF[20] (b) and RD-VPT2 (c) frequencies (vertical lines) superimposed to the HREEL spectra[175] in the 400-1600 cm^{-1} frequency range. [A] CH_2 GLYIp bend.; [B] NH_2CH_2 twist; [C] CH_2 z bend.; [D] CO+CC str.; [E] NH_2CH_2 twist; [F] CN str.; [G] CC str.; [H] NH_2CH_2 twist; [I] OCO bend.; [J] SiO str.; [K,L] strongly delocalised vibrations.

Chapter 4

Environmental effects: theory



In this Chapter some developments of the DFTB methods done in the last years such as the analytical parametrization,[131] the application of the linear response theory (TD-DFTB[134,198]) and the inclusion[132] of the solvation effects using the PCM will be presented. In Section 4.1 the theoretical background of the DFTB method is discussed, focusing on the most relevant aspects of theory, parametrization and implementation. In Section 4.2 the theory and implementation of the DFTB/PCM scheme are illustrated in detail for energies, first and second derivatives, including the contributions to the coupled perturbed self-consistent field (CP-SCF) equations. The TD-DFTB and TD-DFTB/PCM approaches are then presented together in Section 4.3. The DFTB/PCM and TD-DFTB/PCM schemes have been implemented in a localy modified development version of Gaussian code[133] (namely GDV.H21).

4.1 The DFTB method

4.1.1 The Semi Eempirical Tight Binding method

Within the Density Functional Theory (DFT), the ground state energy functional of a closed-shell molecule composed by N electrons and N_A nuclei can be written as[111]

$$\mathcal{E}[n] = \sum_i^N \langle\psi_i| -\frac{\nabla^2}{2} + v + \frac{1}{2}\int d^3\vec{r}\,'\frac{n(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}|\psi_i\rangle + E_{XC}[n] + \frac{1}{2}\sum_{AB}^{N_A}\frac{Z_A Z_B}{R_{AB}} \quad (4.1)$$

where $R_{AB} = |\vec{R_A} - \vec{R_B}|$, $n = n_\alpha + n_\beta$ is the total density, $\{\psi_i,\, i = 1, ..., \mathrm{MO}\}$, are the Kohn-Sham molecular orbitals, E_{XC} is the exchange-correlation energy and v the external potential. It has been demonstrated[113,114] that such an energy functional can be transformed using a reference density (n_0) approximated by a superposition of neutral and spin-unpolarized atomic densities, resulting in[111]

$$\begin{aligned}\mathcal{E}_0[n_0] = &\sum_i^N \langle\psi_i|\,\mathcal{H}_0\,|\psi_i\rangle + E_{XC}[n_0] - \int d^3\vec{r}\quad n_0 \cdot V_{XC}[n_0] \\ &- \frac{1}{2}\int d^3\vec{r} d^3\vec{r}\,' \cdot \frac{n_0(\vec{r})\cdot n_0(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + \frac{1}{2}\sum_{AB}^{N_A}\frac{Z_A Z_B}{R_{AB}}\end{aligned} \quad (4.2)$$

The Hamiltonian $\mathcal{H}_0$ is formally defined as

$$\mathcal{H}_0 = -\frac{\nabla^2}{2} + v + \int d^3\vec{r}\,'\frac{n_0(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + V_{XC}[n_0] \quad (4.3)$$

and is used to find the orbitals $|\psi_i\rangle$, by solving in a non-self-consistent manner a set of Kohn-Sham-like equations of the type

$$\mathcal{H}_0\,|\psi_i\rangle = \varepsilon_i\,|\psi_i\rangle\,. \quad (4.4)$$

Then, the first term of Eq. 4.2 can be written as the sum of the eigenvalues of $\mathcal{H}_0$, the other terms can be approximated[114] by a sum of short range repulsive pairwise potentials U_{AB} and the total energy reads[111,113–115,199,200]

$$E = \sum_i^N \varepsilon_i + \frac{1}{2}\sum_{AB}^{N_A} U_{AB}(R_{AB}). \quad (4.5)$$

This is the energy expression of the so called Semi Empirical Tight Binding (SETB[199,200]) method and in this formalism the first term is usually called the "band structure"

energy (E_{BS}), while the second one is the "repulsive" energy (E_{rep}).[a] The frozen core approximation is also used and only the valence electrons are explicitely treated.

Although the SETB method was successfully used in many cases[201–204] for the computation of termodynamical properties and band structures of solids, two main aspects are particularly critical, one concerning the electrostatic effects (i), the other the treatment of the open-shell systems (ii).

4.1.2 The Self Consistent Charge DFTB method

The first aspect (i) is related to the fact that the SETB method is based on a fixed reference charge density and no charge redistribution between the atoms (or, in terms of spin-orbitals, the orbital relaxation) is allowed. In order to improve the electrostatics the charge redistribution was included by Elstner *et al.*[111] as a second order correction to the total energy. In particular, they proposed to expand Eq. 4.2 in Taylor series about n_0, with respect to the charge density fluctuations (δn). The first derivative is identically null by definition, while the second order term survives, so that Eq. 4.2 can be rewritten as

$$\mathcal{E}_0^{2\mathrm{nd}}[n_0,\delta n] = \mathcal{E}_0[n_0] + \frac{1}{2}\int d^3\vec{r}d^3\vec{r}' \left[\frac{1}{|\vec{r}-\vec{r}'|} + \frac{\delta^2 E_{XC}}{\delta n(\vec{r})\delta n(\vec{r}')}\right]_{n=n_0} \delta n(\vec{r})\delta n(\vec{r}') \quad (4.6)$$

Then, from $\mathcal{E}_0[n_0]$ the energy of the SETB method is obtained, while the other term can be approximated by expanding the density fluctuation in multipoles and truncating at the first order

$$\delta n = \sum_A^{N_A} \Delta q_A f_A(|\vec{r}-\vec{R}_A|). \quad (4.7)$$

The quantities $\Delta q_A = q_A - Z_A$ are the total charges of the atom A, q_A and Z_A being respectively the valence electronic and nuclear charges,[b] f_A are spherically symmetric and non-overlapping functions centered on atom A and representing the radial distribution of the charges. The factor $(4\pi)^{-1}$ originating from the spherical

[a] A more correct formalism for $U_{AB}(R_{AB})$ in Eq. 4.5 should be $U_{\mathrm{type}(A),\mathrm{type}(B)}(R_{AB})$, since for a given couple of atoms A and B, the functional form of U depends on the typology of such atoms. For example, if atom 1 is a Silicon and atom 2 is an Hydrogen $U_{\mathrm{type}(A),\mathrm{type}(B)}(R_{AB}) = U_{Si,H}(|\vec{R_1}-\vec{R_2}|)$. For a binary system composed only by Silicon and Hydrogen atoms the repulsive energy can be written as

$$E_{\mathrm{rep}} = \sum_{\substack{A\in[Silicon]\\ B\in[Hydrogen]}} U_{Si,H}(R_{AB}) + \sum_{\substack{AB\in[Hydrogen]\\ A>B}} U_{H,H}(R_{AB}) + \sum_{\substack{AB\in[Silicon]\\ A>B}} U_{Si,Si}(R_{AB})$$

[b] For example, for a Silicon atom we always have a fix core charge $Z_{Si} = +4$, while the electronic charge is computed from the electronic density and it has a value of -4 for an isolated neutral atom, while it is generally different from -4 when the atom interacts with the neighbours.

harmonic associated to the first multipole is also included in such functions. By inserting Eq. 4.7 in Eq. 4.6 the second order functional can be expressed by a sum over atoms,

$$E = E_{\mathrm{BS}} + E_{\mathrm{rep}} + \frac{1}{2}\sum_{AB}^{N_A} \Delta q_A \gamma_{AB} \Delta q_B \tag{4.8}$$

the kernel being defined as

$$\gamma_{AB} = \int d^3\vec{r} d^3\vec{r}' \left[\frac{1}{|\vec{r}-\vec{r}'|} + \frac{\delta^2 E_{XC}}{\delta n^T(\vec{r})\delta n^T(\vec{r}')}\right]_{n^T=n_0^T} f_A(|\vec{r}-\vec{R}_A|) f_B(|\vec{r}'-\vec{R}_B|). \tag{4.9}$$

The last term of Eq. 4.8 is usually referred to as the "self-consistent-charge" contribution (E_{SCC}). By assuming an exponential decay of the charge distributions and neglecting the derivatives of E_{XC}, the integral in Eq. 4.9 can be evaluated analytically[111]

$$\gamma_{AB} = \frac{1}{R_{AB}} - S(\tau_A, \tau_B, R_{AB}), \tag{4.10}$$

where τ_A and τ_B are atomic quantities connected to the chemical hardness of atoms A and B. The function S is a damping function[c] which ensures that the integral converges to the chemical hardness of atom A when $R_{AB} = 0$, removing the singularity. The method based on the energy expression of Eq. 4.8 is usually called the Self-Consistent-Charge DFTB (SCC-DFTB[111]) or also DFTB for conciseness, and in order to compute such an energy a self-consistent-field (SCF) procedure must be applied being E_{SCC} quadratic in the density (charges).

It is worth noting that the DFTB method can be also derived[205] by writing the DFT energy functional (Eq. 4.1) in atomic orbital basis set and applying the Mulliken approximation to the integrals in the form

$$\chi_\mu \chi_\nu \simeq \frac{1}{2} S_{\mu\nu} \left(|\chi_\mu|^2 + |\chi_\nu|^2\right) \tag{4.12}$$

so that a generic integral with kernel $K(\vec{r})$ can be written as

$$\int d^3\vec{r}\, \chi_\mu(\vec{r})\, K(\vec{r})\, \chi_\nu(\vec{r}) \simeq \frac{1}{2} S_{\mu\nu} \int d^3\vec{r}\, \left[\chi_\mu(\vec{r})\, K(\vec{r})\, \chi_\mu(\vec{r}) + \chi_\nu(\vec{r})\, K(\vec{r})\, \chi_\nu(\vec{r})\right] \tag{4.13}$$

and when μ and ν are located on different atoms, the two-center integrals (and similarly it can be extended also to the three- and four-center integrals) are approximated by an average of one-center integrals weighted by the corresponding Overlap

[c] From Ref.[111]

$$S = e^{-\tau_A R_{AB}} \left[C_{AB}(\tau_A, \tau_B) - \frac{D_{AB}(\tau_A, \tau_B)}{R_{AB}}\right] + e^{-\tau_B R_{AB}} \left[C_{BA}(\tau_A, \tau_B) - \frac{D_{BA}(\tau_A, \tau_B)}{R_{AB}}\right] \tag{4.11}$$

matrix element. The various integrals are then collected into atomic charges and Eq. 4.8 arises. Such an approximation is analogous to the application of the Zero Differential Overlap approximation to the Hartree-Fock energy.[101–106]

We can graphically represent the Mulliken approximation in one dimension by plotting the left and right hand sides of Eq. 4.12 using the two trial functions

$$\chi_\mu(\vec{r}) \to f(x) = e^{-x^2} \quad \text{and} \quad \chi_\nu(\vec{r}) \to g(x) = 2e^{-2x^2} \tag{4.14}$$

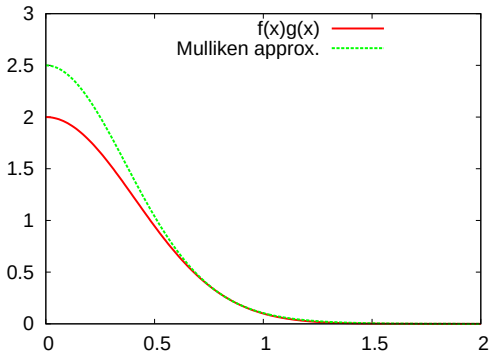


Figure 4.1: Mulliken approximation in one dimension using the trial functions of Eq. 4.14.

In Figure 4.1 it can be observed that for two gaussian-like charge distributions the Mulliken approximation is valid for large interatomic distances, where the Overlap between χ_μ and χ_ν is small. But this condition is just equivalent to say that the reference density (n_0) – defined as a superposition of atomic densities – is a good approximation of the real molecular density. Thus, for large interatomic distances the SETB functional is a good approximation of the ground state DFT energy (at least from a theoretical point of view, diregarding the parametrization schemes), while for smaller distances the second order contribution depending on the charge fluctuations provides small corrections to the Mulliken approximation.

4.1.3 The spin-dependent SCC-DFTB method

A similar approach has been applied[206–208] also for the treatment of the spin-polarization effects (ii), as originally proposed by Köler *et al.*[206] In particular, defining the spin-density as $n^S = n_\alpha - n_\beta$, and specifying the superscript to the total density

($n^T = n_\alpha + n_\beta$), the spin-dependent couterpart of Eq. 4.1 can be written as[206]

$$\begin{aligned}\mathcal{E}[n^T, n^S] = \sum_{\substack{i \\ \sigma=\alpha,\beta}}^{N} n_{i\sigma} \left\langle \psi_{i\sigma} \right| - \frac{\nabla^2}{2} + v + \frac{1}{2}\int d^3\vec{r}\,' \frac{n^T(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \left| \psi_{i\sigma} \right\rangle \\ + E_{XC}[n^T, n^S] + \frac{1}{2}\sum_{AB}^{N_A} \frac{Z_A Z_B}{R_{AB}}\end{aligned} \tag{4.15}$$

This functional can be approximated by using the reference total density employed above (which is now labelled as n_0^T), and by generalizing with the introduction of the reference spin density (n_0^S). Being n_0^T a superposition spin-unpolarized (and neutral) atomic densities, the reference spin-density is zero and Eq. 4.2 can be considered as an approximation of Eq. 4.15 at n_0^T and $n_0^S = 0$. In order to include the spin-polarization as a second-order effect, Eq. 4.2 is expanded in both the total (δn^T) and the spin (δn^S) density fluctuations, about n_0^T and $n_0^S = 0$, obtaining

$$\begin{aligned}\mathcal{E}_0^{\text{2nd}}[n_0^T, n_0^S, \delta n^T, \delta n^S] =& \mathcal{E}_0[n_0^T, n_0^S] \\ &+ \frac{1}{2}\int d^3\vec{r} d^3\vec{r}\,' \left[\frac{1}{|\vec{r} - \vec{r}\,'|} + \frac{\delta^2 E_{XC}}{\delta n^T(\vec{r})\delta n^T(\vec{r}\,')}\right]_{n^T = n_0^T} \delta n^T(\vec{r})\delta n^T(\vec{r}\,') \\ &+ \frac{1}{2}\int d^3\vec{r} d^3\vec{r}\,' \left[\frac{\delta^2 E_{XC}}{\delta n^S(\vec{r})\delta n^S(\vec{r}\,')}\right]_{n^S = 0} \delta n^S(\vec{r})\delta n^S(\vec{r}\,')\end{aligned} \tag{4.16}$$

More details about this expansion can be found in the original works.[111,114,206] From $\mathcal{E}_0[n_0^T, n_0^S]$ the energy of the SETB method is obtained, while from the second order term in the total density fluctuations the SCC-DFTB energy arises. The third term is approximated by expanding the spin density fluctuations in multipoles and by truncating the series at the first order

$$\delta n^S = \sum_{\substack{Al \\ l \in A}} p_{Al} f_{Al}(|\vec{r} - \vec{R}_A|). \tag{4.17}$$

The index A runs over all the atoms composing the system, l runs over the angular momenta (s, p, d, ...) of the electrons of atom A, p_{Al} is the charge associated to the valence electrons on atom A with angular momentum l, while f_{Al} are analogous to the f_A functions introduced above. By inserting Eq. 4.17 in Eq. 4.16 the functional depending on the spin density fluctuation can be expressed by a sum over atoms,

$$E_{\text{spin}} = \frac{1}{2}\sum_{\substack{All' \\ ll' \in A}} p_{Al} W_{All'} p_{Al'} \tag{4.18}$$

the kernel being defined as

$$W_{ll'}^A = \int d^3\vec{r} d^3\vec{r}\,' \left[\frac{\delta^2 E_{XC}}{\delta n^S(\vec{r})\delta n^S(\vec{r}\,')}\right]_{n^S = 0} f_{Al}(|\vec{r} - \vec{R}_A|) f_{Al'}(|\vec{r}\,' - \vec{R}_A|). \tag{4.19}$$

The integral of Eq. 4.19 is an atomic quantity (depends only on index A) and can be evaluated[206] using the Janak theorem.[209] Including also the spin polarization[206,207] the final energy can be written as

$$\begin{aligned} E =& E_{\mathrm{BS}} + E_{\mathrm{rep}} + E_{\mathrm{SCC}} + E_{\mathrm{spin}} \\ =& \sum_{i\sigma}^{N} \varepsilon_{i\sigma} + \frac{1}{2}\sum_{AB} U_{AB} + \frac{1}{2}\sum_{AB} \gamma_{AB}\Delta q_A \Delta q_B + \frac{1}{2}\sum_{All'} p_{lA} W_{ll'A} p_{l'A} \end{aligned} \tag{4.20}$$

Due to the presence of E_{spin}, at the end of the SCF the α and β charge densities and eigenvalues are in principle different.

4.1.4 Parameters of the DFTB

As the other semi-empirical methods, the DFTB includes a set of user-given parameters in order to practically perform the calculations, and the accuracy of the results is strongly dependent on the employed parametrization scheme. The DFTB total energy expression is composed by four functionals, namely E_{BS}, E_{rep}, E_{SCC} and E_{spin}. The parameters entering in E_{SCC} and E_{spin} are atomic quantities and have been already discussed in the previous section. In particular, one τ_A per atom needs to be specified in the functional E_{SCC} and can be computed from the atomic chemical hardness. The functional E_{spin} depends on the $W_{All'}$ parameters, which can be evaluated using the Janak theorem.[209] For an sp^3 atom three parameters need to be defined, namely W_{Ass}, W_{Asp} and W_{App}, becoming six when also the d charge distribution is considered (W_{Asd}, W_{Apd} and W_{Add}).

In order to introduce the parameters of the band structure energy, it is convenient to rewrite E_{BS} in atomic orbital basis set $\{\chi_\mu\}, \mu = 1, ..., \mathrm{AO}$

$$E_{\mathrm{BS}} = \sum_{i}^{N} \varepsilon_i = \sum_{i}^{N} \langle\psi_i|\hat{H}^0|\psi_i\rangle = \sum_{\mu\nu}^{AO} P^T_{\mu\nu}\langle\chi_\mu|\hat{H}^0|\chi_\nu\rangle = \sum_{\mu\nu}^{AO} P^T_{\mu\nu} h^0_{\mu\nu} \tag{4.21}$$

so that the corresponding Fock matrix elements are just

$$F_{\mu\nu} = h^0_{\mu\nu} \tag{4.22}$$

and the eigenvalue equations in matrix representation become

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \tag{4.23}$$

where $\mathbf{S}$ is the Overlap matrix, $\boldsymbol{\varepsilon}$ is the set of eigenvalues, and $\mathbf{C}$ is the set of coefficients defining the density matrix. It is worth noting that Eq. 4.22 corresponds to the Fock matrix element of the SETB method, which is clearly non-self-consistent, $F_{\mu\nu}$ not depending on the density matrix. In the first tight binding schemes the Overlap matrix was usually set to the identity in order to simplify the calculations but subsequently it has been shown that a non-orthogonal scheme strongly

improves the transferability of the method.[202] Thus, a set of parameters is required in order to compute the matrix elements of **F** and **S**. In a two-center Slater-Koster sp^3 parametrization scheme there exists four Fock interactions (h: $h^{pp-\sigma}$, $h^{pp-\pi}$, $h^{sp-\sigma}$, $h^{ss-\sigma}$) and four Overlap interactions (s: $s^{pp-\sigma}$, $s^{pp-\pi}$, $s^{sp-\sigma}$, $s^{ss-\sigma}$). Such a parametrization is effective for the elements of the first and second period, while for heavier elements the inclusion of the d orbitals can be necessary. In a two-center sp^3d^5 parametrization scheme there are ten interactions for both the Fock and Overlap matrix elements, since also the dd$-\sigma$, dd$-\pi$, dd$-\delta$, pd$-\sigma$, pd$-\pi$, sd$-\sigma$ must be included. The usual procedure[199] is to precompute the values of the integrals for dimers in a given range of interatomic distances and to store them in proper tables. When the Fock and Overlap matrices are built for the real calculations such values are read and the corresponding matrix elements are obtained.[d]

Once $E_{\rm BS}$ is defined, the pair potentials U_{AB} of the repulsive energy can be obtained by fitting the difference between the total energy of a reference method and $E_{\rm BS}$. In some parametrization schemes[119] also the gradients and the Hessians have been used in order to improve the accuracy of the vibrational properties. Two functional forms are usually employed for the repulsive potentials. The first representation is as a polynomial function of the interatomic distance (R)

$$U_{AB}(R) = \sum_{n=2}^{9} c_n (R_{\rm cut} - R)^n \tag{4.24}$$

where $R_{\rm cut}$ is the scalar user-given cutoff distance of the repulsion. The other form is rather complex. In particular, the full distance range $[0, R_{\rm cut}]$ is divided into a set of smaller intervals,

$$[0, R_{\rm cut}] = [R_0, R_1] \cup [R_1, R_2] \cup [R_2, R_3] ... \cup [R_{K-1}, R_K] \tag{4.25}$$

such that $R_0 = 0$, $R_K = R_{\rm cut}$. In the first interval, i.e. the short-range region, an exponential decay is assumed

$$e^{a_1 R + a_2} + a_3 \qquad R \in [R_0, R_1], \tag{4.26}$$

where a_1, a_2, a_3 are user-given parameters; in the range $[R_1, R_{K-1}]$, the repulsion is fitted with splines of the type

$$\sum_{\substack{n=0 \\ R \in [R_J, R_{J+1}]}}^{3} c_n (R - R_J)^n \qquad J \in [1, K-2] \tag{4.27}$$

where the set of user-given parameters is $\{c_n\}, n = 0, 1, 2, 3$. Finally, in the last interval $[R_{K-1}, R_K]$ the sum of Eq. 4.27 includes up to the fifth order terms, in

[d] Please note that since the h and s parameters are obtained from calculations on dimers, direction cosines need to be used in order to consider the right orientation of the pairs of atoms in the molecule when the $h^0_{\mu\nu}$ matrix elements are computed.

order to provide a smooth decay to the cutoff value. Thus, for a given pair of atoms A and B, the repulsive potential reads

$$U_{AB}(R) = \begin{cases} e^{a_1 R + a_2} + a_3 & R \in [R_0, R_1] \\ \sum_{n=0}^{3} c_n (R - R_J)^n & R \in [R_J, R_{J+1}], J \in [1, K-2] \\ \sum_{n=0}^{5} c_n (R - R_{K-1})^n & R \in [R_{K-1}, R_K]. \end{cases} \tag{4.28}$$

This functional form makes the potential function very flexible and, in principle, the transferability can be increased by increasing the number of the sub-intervals.

One of the problems encountered with the DFTB method is that the Slater-Koster tabulated parameters and the functional form of the repulsive potentials can be affected by discontinuities, which can be annoying when the analytic derivatives of the energy are computed (see Section 4.1.5). In particular, since the values of h and s are tabulated, the analytical derivatives of the curves are not availalble and they can be only evaluated numerically from the incremental ratio between two consecutive points. Also the repulsive potentials are affected by problems connnected with the discontinuities, since at the borders between two consecutive intervals the derivatives from right and left can be different. Then, in order to ensure that the analytical derivatives of the parameter related to E_{BS} and E_{rep} always exist, a fully analytical version of the DFTB method (usually referred to as DFTBA[112,131,134]) has been developed and implemented in the GAUSSIAN code.[133] The basic idea is to express the Slater-Koster parameters as explicit functions of the interatomic distance, and the repulsive potentials as continuous – rather than piecewise continuous – functions. In particular, for a given pair of atoms all the h, s and U_{AB} parameter curves are represented by analytical functions of R, in the range $R \in [0, R_{\mathrm{cut}}]$, using a sum of monocentric exponentials

$$f(R) = \sum_{n=1}^{10} c_n e^{-\alpha \beta^n R}. \tag{4.29}$$

In order to ensure that all the interactions go smoothly to zero at R_{cut}, a damping function $P(R)$ is applied in the interval $[R_{\mathrm{dump}}, R_{\mathrm{cut}}]$, where R_{dump} is a user-given parameter, so that the effective functional form for all the curves $h(R)$, $s(R)$ and $U_{AB}(R)$ is

$$F(R) = \begin{cases} f(R) & R \in [0, R_{\mathrm{dump}}[\\ f(R)P(R) & R \in [R_{\mathrm{dump}}, R_{\mathrm{cut}}] \end{cases} \tag{4.30}$$

The function $P(R)$

$$\begin{aligned} P(R) &= R_s^3(10 - 15R_s + 6R_s^2) \\ R_s &= \frac{R_{\mathrm{cut}} - R}{R_{\mathrm{cut}} - R_{\mathrm{dump}}} \end{aligned} \tag{4.31}$$

is chosen in order to ensure that

$$\begin{aligned} F(R_{\text{dump}}) &= f(R_{\text{dump}}) \\ F'(R_{\text{dump}}) &= f'(R_{\text{dump}}) \\ F''(R_{\text{dump}}) &= f''(R_{\text{dump}}) \\ F(R_{\text{cut}}) &= F'(R_{\text{cut}}) = F''(R_{\text{cut}}) = 0. \end{aligned} \tag{4.32}$$

Thus, $F(R)$ is completely determined by the set of coefficients $\{\alpha, \beta, c_n, n = 1, 2, ...10\}$, R_{cut} and R_{dump}.[e] In general, the parameters of E_{BS} and E_{rep} are by far the most important part of the parametrization, being E_{SCC} and E_{spin} only second order corrections to the total energy. Many – classical[119,121,210,211] and analytical[131,208] – parametrization schemes have been developed in order to compute specific chemico-physical properties, such as band structures of solids,[210,211] vibrational frequencies of organic molecules,[119] excitation properties of nanoparticles,[131] and also systematic procedures for the parametrization of the repulsive potentials over an extended set of molecules have been proposed.[121]

4.1.5 Implementation in GAUSSIAN and derivatives

The implementation in the GAUSSIAN code[133] is based on the atomic orbital (AO) expansion of the DFTB energy of Eq. 4.20. In the previous section it was shown that E_{BS} is a monoelectronic functional linearly dependent on the density matrix, while the repulsive energy does not depend on the density matrix. In order to write $F_{\mu\nu}$ for the DFTB method the total and spin charges must be expanded in atomic orbital basis set. Usually, the Mulliken charges are employed for the evaluation of E_{SCC} and E_{spin}, although different choices such as the CM3[128–130] have been also proposed. The total Mulliken charge for the atom A can be written as

$$\Delta q_A = q_A - Z_A = \sum_{\substack{\mu\nu \\ \mu\in A}}^{\text{AO}} \frac{1}{2}(P^T_{\mu\nu} S_{\mu\nu} + P^T_{\nu\mu} S_{\nu\mu}) - Z_A \tag{4.33}$$

where $\mathbf{P}^T = \mathbf{P}^\alpha + \mathbf{P}^\beta$ is the total density matrix. By substituting Eq. 4.33 in Eq. 4.20, the E_{SCC} functional can be splitted in zero- ($E^{(0)}_{\text{SCC}}$), mono- ($E^{(1)}_{\text{SCC}}$) and

[e] The value currently employed in the GAUSSIAN code[133] for R_{dump} is $0.85 \cdot R_{\text{cut}}$.

bi-electronic ($E^{(2)}_{\mathrm{SCC}}$) contributions

$$\begin{aligned}
E^{(0)}_{\mathrm{SCC}} =& \frac{1}{2}\sum_{AB} Z_A \gamma_{AB} Z_B \\
E^{(1)}_{\mathrm{SCC}} =& -\sum_{AB} q_A \gamma_{AB} Z_B \\
=& -\sum_{A} \frac{1}{2} \sum_{\substack{\mu\nu \\ \mu\in A}} (P^T_{\mu\nu} S_{\mu\nu} + P^T_{\nu\mu} S_{\nu\mu}) \sum_B \gamma_{AB} Z_B \\
E^{(2)}_{\mathrm{SCC}} =& -\frac{1}{2}\sum_{AB} q_A \gamma_{AB} q_B \\
=& -\sum_{A} \frac{1}{2} \sum_{\substack{\mu\nu \\ \mu\in A}} (P^T_{\mu\nu} S_{\mu\nu} + P^T_{\nu\mu} S_{\nu\mu}) \sum_B \gamma_{AB} q_B
\end{aligned} \tag{4.34}$$

The double sum over A and over $\mu \in A$ is just a sum over the complete AO basis set for the μ index and we can establish the equality

$$\sum_A \sum_{\substack{\mu\nu \\ \mu\in A}} = \sum_{\mu\nu} \tag{4.35}$$

In this formalism, the A index of γ_{AB} is always referred to the atom on which the χ_μ atomic function is located so that will be indicated as (μ). The mono- and bielectronic functionals can be then rewritten as

$$\begin{aligned}
E^{(1)}_{\mathrm{SCC}} =& -\frac{1}{2}\sum_{\mu\nu} P^T_{\mu\nu} S_{\mu\nu} \sum_B \left[\gamma_{(\mu)B} + \gamma_{(\nu)B}\right] Z_B \\
E^{(2)}_{\mathrm{SCC}} =& \frac{1}{4}\sum_{\mu\nu} P^T_{\mu\nu} S_{\mu\nu} \sum_B \left[\gamma_{(\mu)B} + \gamma_{(\nu)B}\right] q_B
\end{aligned} \tag{4.36}$$

Writing the p_{Al} charges in terms of the spin density matrix $\mathbf{P}^S = \mathbf{P}^\alpha - \mathbf{P}^\beta$

$$p_{Al} = \sum_{\substack{\mu\nu \\ \mu\in(A,l)}} P^S_{\mu\nu} S_{\mu\nu} \tag{4.37}$$

and substituting in Eq. 4.20, the E_{spin} functional is written as a bielectronic contribution

$$\begin{aligned}
E_{\mathrm{spin}} =& \frac{1}{2}\sum_{All'} p_{Al} W_{All'} p_{Al'} \\
=& \frac{1}{2}\sum_{Al} \sum_{\substack{\mu\nu \\ \mu\in(Al)}} P^S_{\mu\nu} S_{\mu\nu} \sum_{l'} W_{All'} p_{Al'} \\
=& \frac{1}{4}\sum_{\mu\nu} P^S_{\mu\nu} S_{\mu\nu} \sum_{l'} \left[W_{(\mu)[\mu]l'} p_{(\mu)l'} + W_{(\nu)[\nu]l'} p_{(\nu)l'}\right]
\end{aligned} \tag{4.38}$$

where the index in round parenthesis refers to the atom on which the atomic basis function is centered while the index in square parenthesis refers to the angular momentum of the basis function.

The DFTB total energy is then composed by a mono-electronic contribution including the linear functionals of the density matrix (E_{BS} and $E^{(1)}_{\mathrm{SCC}}$), a bielectronic contribution including the quadratic functionals of the density matrix (E_{spin} and $E^{(2)}_{\mathrm{SCC}}$) and a zeroth order term V, including E_{rep} and $E^{(0)}_{\mathrm{SCC}}$ which do not depend on the density matrix. Finally, the DFTB total energy can be written in terms of $\mathbf{P}^{\alpha}$ and $\mathbf{P}^{\beta}$

$$E = \left\langle \mathbf{P}^{\alpha} \left(\mathbf{h} + \frac{1}{2}\mathbf{G}^{\alpha} \right) \right\rangle + \left\langle \mathbf{P}^{\beta} \left(\mathbf{h} + \frac{1}{2}\mathbf{G}^{\beta} \right) \right\rangle + V \tag{4.39}$$

where

$$\begin{aligned} h_{\mu\nu} &= h^{0}_{\mu\nu} - \frac{1}{2} S_{\mu\nu} \sum_{B} \left[\gamma_{(\mu)B} + \gamma_{(\nu)B} \right] Z_B \\ G^{\alpha/\beta}_{\mu\nu} &= \frac{1}{2} S_{\mu\nu} \left[\sum_{B} \left[\gamma_{(\mu)B} + \gamma_{(\nu)B} \right] q_B \pm \sum_{l'} \left[W_{(\mu)[\mu]l'} p_{(\mu)l'} + W_{(\nu)[\nu]l'} p_{(\nu)l'} \right] \right] \\ V &= \frac{1}{2} \sum_{AB} \left(U_{AB} + Z_A \gamma_{AB} Z_B \right) \end{aligned} \tag{4.40}$$

and $\langle \quad \rangle$ is the matrix trace operation. By using the following definition of the Fock matrix

$$\mathbf{F}^{\alpha/\beta} = \mathbf{h} + \mathbf{G}^{\alpha/\beta} \tag{4.41}$$

the total energy reads[112]

$$E = \frac{1}{2} \left\langle \mathbf{P}^{\alpha} \left(\mathbf{h} + \mathbf{F}^{\alpha} \right) \right\rangle + \frac{1}{2} \left\langle \mathbf{P}^{\beta} \left(\mathbf{h} + \mathbf{F}^{\beta} \right) \right\rangle + V. \tag{4.42}$$

In Appendix B the total and spin charges, Fock matrix elements and total energies for some simple atomic systems are reported in order to illustrate how Eq. 4.33, 4.37, 4.41 and 4.42 are practically applied to the calculation of the DFTB quantities.

Within this formalism the derivation of the analytic derivatives of the DFTB energy with respect to an arbitrary perturbation x, resembles the Hartree-Fock one.[212] In case of unrestricted calculations the α and β blocks of Eq. 4.42 can be treated separately, while in case of restricted calculations, they are identical.[212] So the derivation can be simplified by removing the spin index and the first derivative of Eq. 4.42 can be calculated as

$$\begin{aligned} E^{x} =& \frac{\partial E}{\partial x} + \frac{\partial E}{\partial \mathbf{P}} \frac{\partial \mathbf{P}}{dx} \\ =& \langle \mathbf{P}\mathbf{h}^{x} \rangle + \frac{1}{2} \langle \mathbf{P}\mathbf{G}^{(x)} \rangle - \langle \mathbf{W}\mathbf{S}^{x} \rangle + V^{x} \end{aligned} \tag{4.43}$$

where $\mathbf{W} = \mathbf{PFP}$ is the energy weighted density matrix[212] and the derivatives of the different terms can be directly obtained from Eq. 4.40. The $\mathbf{h}^x$ matrix includes all the contributions arising from the derivatives of the monoelectronic functionals, while the $\mathbf{G}^{(x)}$ matrix includes all the derivatives of the bielectronic functionals which do not depend on $\mathbf{P}^x$. The evaluation of $\mathbf{P}^x$ can be avoided in the first derivative computations since the corresponding terms can be expressed by $\mathbf{WS}^x$. So, when the derivatives of the charges (Δq_A^x and $p_{All'}^x$) are performed, the terms depending on $\mathbf{P}^x$ are included in the $\mathbf{WS}^x$ gradient contribution, while the terms depending on $\mathbf{S}^x$ are included in the $\mathbf{G}^{(x)}$ one. In general, the derivatives of all the parameters need to be computed, as was outlined in Section 4.1.4.

The analytical second derivatives of the DFTB energy are computed as the derivatives of Eq. 4.43 with respect to an arbitrary perturbation y. Such a derivation would not add any new significant point to the present work, so that we address the interested reader to the original articles.[112,116] We only recall that for the second derivatives the explicit evaluation of $\mathbf{P}^x$ is required, so that the CP-SCF equations need to be solved by explicitely specifying the nature of the pertubations x and y. In case of the derivatives with respect to the nuclear displacements ($\mathcal{Q}$) and to an applied electric field ($\mathcal{F}$) the implementation in the GAUSSIAN code[133] is based on the definition of a unique matrix of the type

$$\begin{pmatrix} \frac{\partial^2 E}{\partial \mathcal{F}^2} & \frac{\partial^2 E}{\partial \mathcal{Q} \partial \mathcal{F}} \\ \\ \frac{\partial^2 E}{\partial \mathcal{F} \partial \mathcal{Q}} & \frac{\partial^2 E}{\partial \mathcal{Q}^2} \end{pmatrix} \tag{4.44}$$

where the upper-left block is the polarizability tensor (3×3) the lower-right block is the molecular Hessian ($3N_A - 6 \times 3N_A - 6$[f]) from which the harmonic frequencies are computed, and the off-diagonal blocks are related to the harmonic Infrared intensities.

[f] $3N_A - 5 \times 3N_A - 5$ for linear molecules.

4.2 The DFTB/PCM scheme

4.2.1 Brief review of the PCM formalism

The interaction energy between the solute charge density $n(\vec{r})$ (both nuclear and electronic) and the apparent solvent charge distribution $\sigma(\vec{s})$ is

$$E_{\text{int}} = \int_{\Gamma} d\vec{s} \int d\vec{r} \frac{n(\vec{r})\sigma(\vec{s})}{|\vec{r} - \vec{s}|} = \int_{\Gamma} d\vec{s} V(\vec{s})\sigma(\vec{s}) \tag{4.45}$$

where $\vec{r} \in \mathbb{R}^3$ and $\vec{s}$ is a point on the surface Γ that corresponds to the solute-solvent interface which separates the solute from the polarizable medium that represents the solvent. $V(\vec{s})$ is the molecular electrostatic potential generated by the charge distribution $n(\vec{r})$ at $\vec{s}$. Following the formalism of Barone et al.,[213] the free energy of the interacting solute-solvent system is

$$\mathcal{G} = E^0[n] + V_{NN} + \frac{1}{2} E_{\text{int}} \tag{4.46}$$

where $E^0[n]$ is the solute's electronic energy functional and V_{NN} is the *nuclear* electrostatic interaction, which does not depend on the electronic density. The functional forms of $E^0[n]$ and V_{NN} depend on the specific methodology employed for the calculation of the solute energy and charge density. In DFT calculations, the charge density can be expressed in a basis set of atomic orbitals (AO) as

$$n(\vec{r}) = \sum_{\mu\nu}^{N_B} P_{\mu\nu}\chi_\mu(\vec{r})\chi_\nu(\vec{r}) - \sum_{A}^{N_A} Z_A \delta(\vec{r} - \vec{R}_A) \tag{4.47}$$

where μ and ν indexes run over the functions of the AO basis set $\{\chi_\mu\}, \mu = 1...N_B$ and N_A is the number of atoms. If the nuclear charge density is treated as a distribution of point charges $\{Z_A\}$, the electrostatic potential at point $\vec{r}$ associated to $n(\vec{r})$ is

$$V(\vec{r}) = \sum_{\mu\nu} P_{\mu\nu} \int d\vec{r}\,' \frac{\chi_\mu(\vec{r}\,')\chi_\nu(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} - \sum_{A} \frac{Z_A}{|\vec{r} - \vec{R}_A|} \tag{4.48}$$

In the practical implementation of the PCM model, the surface Γ is discretized into a certain number n_g of surface elements and the solvent charge distribution $\sigma(\vec{s})$ is expressed as a function of $\vec{r} \in \mathbb{R}^3$

$$\sigma(\vec{r}) = \sum_{i}^{n_g} \frac{q_i}{a_i} \phi_i(\vec{r}; \vec{s_i}). \tag{4.49}$$

The quantities a_i and $\vec{s_i}$ are the geometrical parameters that define the discretization of Γ, and correspond to the area and the coordinates of the center of the i-th surface element. The apparent solvent charge distribution is represented by a collection

of q_i point charges and $\phi_i(\vec{r};\vec{s_i})$ gaussian functions associated with each surface element. Having introduced such a discretization of the solute-solvent interface, the interaction energy can be written as a sum rather than an integral,

$$E_{\text{int}} = \mathbf{V}^{\dagger}\mathbf{q} = \sum_i V_i q_i. \tag{4.50}$$

The $\mathbf{q}$ vector contains the charges q_i, which expand $\sigma(\vec{s})$, while the elements of $\mathbf{V}$ are defined given the definition of the charge density $n(\vec{r})$ and of the $\phi_i(\vec{r};\vec{s})$ functions

$$V_i = \sum_{\mu\nu} P_{\mu\nu}\langle i|\mu\nu\rangle - \sum_A Z_A\langle i|A\rangle \tag{4.51}$$

where

$$\begin{aligned}\langle i|A\rangle &= \int d\vec{r}\int d\vec{r}\,'\frac{\phi_i(\vec{r};\vec{s}_i)\delta(\vec{r}\,' - \vec{R}_A)}{|\vec{r}-\vec{r}\,'|}\\ \langle i|\mu\nu\rangle &= \int d\vec{r}\int d\vec{r}\,'\frac{\phi_i(\vec{r};\vec{s}_i)\chi_\mu(\vec{r}\,')\chi_\nu(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\end{aligned} \tag{4.52}$$

At each cycle of the self-consistent field (SCF) procedure, $\mathbf{V}$ is computed from the current density matrix, and the vector of charges $\mathbf{q}$ is obtained solving the PCM equations

$$\mathbf{q} = \mathbf{Q}\mathbf{V}. \tag{4.53}$$

The matrix $\mathbf{Q}$ actually corresponds to the inverse matrix of the PCM linear system and it depends on the geometry of the solute, the details of the discretization of the cavity and the particular choice of the PCM model (e.g. D-PCM,[94] C-PCM,[92,93] IEF-PCM[95–97]). The $\mathbf{Q}$ matrix is in general not symmetric and since the energy in Eq. 4.50 actually depends quadratically on the density matrix, the polarization weights ($\mathbf{w}$)

$$\mathbf{w} = \left(\frac{\mathbf{Q}+\mathbf{Q}^{\dagger}}{2}\right)\mathbf{V} \tag{4.54}$$

and not the charges[213] are used in the PCM contribution ($v^{PCM}_{\mu\nu}$) that is added to the gas-phase Fock matrix ($F^0_{\mu\nu}$)

$$\mathbf{F} = \mathbf{F}^0 + \mathbf{v}^{PCM}. \tag{4.55}$$

The PCM operator is defined as

$$v^{PCM}(\vec{r}) = \int_\Gamma d\vec{s}\frac{\sigma(\vec{s})}{|\vec{r}-\vec{s}|} \tag{4.56}$$

and it is discretized as

$$v^{PCM}_{\mu\nu} = \sum_i w_i\langle i|\mu\nu\rangle. \tag{4.57}$$

The interested reader can find more details in the literature about the PCM discretization,[214] the definition of different cavity models,[214,215] the practical implementation of the method,[213] and the formulation of analytical derivatives[213,214,216,217] of the solute energy including the PCM contribution.

4.2.2 Evaluation of electrostatic properties using DFTB

In the DFTB method, the repulsive energy functional (E_{rep}) mixes together the electronic and the nuclear energy contributions (see the previous section) so Eq. 4.46 should be written as

$$\begin{aligned} \mathcal{G} =& E^0[n] + E_{\text{rep}} + \frac{1}{2}E_{\text{int}} \\ E_{\text{rep}} =& \frac{1}{2}\sum_{AB}(V_{AB} + Z_A\gamma_{AB}Z_B). \end{aligned} \tag{4.58}$$

In order to evaluate E_{int}, we recall that the DFTB method is derived from the standard DFT,[111,113–115] and it uses a minimal basis set within the frozen core approximation. Therefore, the charge density of a molecule can be written as in Eq. 4.47, where the AO basis functions χ_μ are those of the minimal valence basis set and the Z_A are the nuclear valence charges. Within the DFTB approach, the atomic basis functions $\{\chi_\mu\}$ are used only as a formal tool for the evaluation of the atomic charges, so that a parametrization scheme is required for the evaluation of the integrals $\langle i|\mu\nu\rangle$, between the solute atomic charges and the basis functions of the apparent solvent charge. However, we want to avoid the introduction of new empirical parameters specifically developed for the integrals required by the PCM model. In fact, such a parametrization effort would be time-consuming and its success could be limited if any inconsistency with the existing Fock and overlap parameters arose. For these reasons, we prefer a simpler and more pragmatic approach for the treatment of the solute-solvent electrostatic interaction, which is able to provide a robust implementation of the PCM model for DFTB, without the introduction of new and specifically optimized parameters.

We start by assuming that the total charge density of a molecule can be expressed, consistently with the DFTB formalism, as the sum of atomic densities $n_A(\vec{r} - \vec{R_A})$

$$\tilde{n}(\vec{r}) = \sum_A n_A(\vec{r} - \vec{R_A}) \tag{4.59}$$

Indeed, this is one of the basic ideas behind the DFTB method and it was already encountered in the previous section. It was originally proposed by Harris et al.[113] in order to approximate the DFT calculations for weakly interacting fragments, and it was subsequently used by Foulkes et al.[114] for the formal derivation of the DFTB. The same approach was also followed by Elstner et al.[111] in order to expand the fluctuations of the total density in atomic contributions, in the derivation of the SCC-DFTB method.[111,112,115] Then, using the formalism of Elstner et al.,[111] each of the n_A densities can be expanded in terms of radial F^A_{lm} and angular Y_{lm} functions with coefficients K^A_{lm}

$$n_A(\vec{r} - \vec{R}_A) = \sum_{l,m} K^A_{lm} F^A_{lm}(|\vec{r} - \vec{R}_A|) Y_{lm}\left(\frac{\vec{r} - \vec{R}_A}{|\vec{r} - \vec{R}_A|}\right). \tag{4.60}$$

If we choose Y_{lm} to be the spherical harmonics and we truncate the expansion to the first term we obtain

$$n_A(\vec{r} - \vec{R}_A) \simeq K_{00}^A F_{00}^A(|\vec{r} - \vec{R}_A|)Y_{00} \tag{4.61}$$

where K_{00}^A is the first charge momentum, which can be computed as the Mulliken atomic charges, while the spherical harmonic with $l = m = 0$ is just $Y_{00} = \frac{1}{\sqrt{4\pi}}$ and we choose the function $F_{00}^A = Y_{00}^{-1}\delta(\vec{r} - R_A)$ as the radial part of the atomic densities.

Introducing these approximations, the total DFTB charge density $\tilde{n}(\vec{r})$ becomes

$$n(\vec{r}) \simeq \tilde{n}(\vec{r}) = \sum_A \Delta q_A \delta(\vec{r} - \vec{R}_A) \tag{4.62}$$

where the quantities Δq_A for symmetric $\mathbf{P}$ and $\mathbf{S}$ matrices are given by

$$\Delta q_A = \sum_{\substack{\mu\nu \\ \mu\in A}} \frac{1}{2}(P_{\mu\nu}S_{\mu\nu} + P_{\nu\mu}S_{\nu\mu}) - Z_A. \tag{4.63}$$

Note that Eq. 4.62 is just an approximation for the *exact* DFTB charge density of Eq. 4.47 at large distances from the sources (i.e. the positions of the nuclei), where the electronic density also appears as a collection of point-charges. The corresponding approximate molecular electrostatic potential $\tilde{V}(\vec{r})$ associated to $\tilde{n}(\vec{r})$ is readily evaluated as

$$\tilde{V}(\vec{r}) = \int d\vec{r}\,' \frac{\tilde{n}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} = \sum_A \frac{\Delta q_A}{|\vec{r} - \vec{R_A}|} = \sum_A \sum_{\substack{\mu\nu \\ \mu\in A}} \frac{1}{2} \frac{(P_{\mu\nu}S_{\mu\nu} + P_{\nu\mu}S_{\nu\mu})}{|\vec{r} - \vec{R_A}|} - \sum_A \frac{Z_A}{|\vec{r} - \vec{R_A}|}. \tag{4.64}$$

Using Eq. 4.35 and recalling from Section 4.1.5 that the notation (μ) indicates the atom where χ_μ is located, the approximate electrostatic potential can also be written as follows,

$$\tilde{V}(\vec{r}) = \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}S_{\mu\nu}\left(\frac{1}{|\vec{r} - \vec{R_{(\mu)}}|} + \frac{1}{|\vec{r} - \vec{R_{(\nu)}}|}\right) - \sum_A \frac{Z_A}{|\vec{r} - \vec{R_A}|} \tag{4.65}$$

For each pair of basis functions χ_μ and χ_ν we can directly compare the integrals of Eq. 4.48 with their approximation in Eq. 4.65. The resulting potential will be close to the *exact* ones as long as the approximation

$$\int d\vec{r}\,' \frac{\chi_\mu(\vec{r}\,')\chi_\nu(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \simeq \frac{1}{2}S_{\mu\nu}\left(\frac{1}{|\vec{r} - \vec{R_{(\mu)}}|} + \frac{1}{|\vec{r} - \vec{R_{(\nu)}}|}\right) \tag{4.66}$$

is acceptable. In particular, the equality between the left and right hand side of the above equation holds exactly when $\vec{r} \gg \vec{r}\,'$ (see Appendix C). This corresponds

to the well known fact that the potential of a distribution of point charges shows the expected $\vec{r}^{-1}$ behavior when it is evaluated far from the location of the charges, while at shorter distances this approximation is no longer valid. If we use such an approximate potential to compute the PCM charges, we expect that for solute-solvent interfaces far enough from the solute (i.e. large solute cavities) it will produce the same result as the *exact* DFTB potential. Furthermore, if the approximation in Eq. 4.66 is equally valid for all the few pairs of AO minimal valence basis functions, we expect the approximate DFTB electrostatic potential to be close to that of a true DFT method within the same minimal basis. At this point all the physics required for the description of the electrostatic solute-solvent interaction within the DFTB approach has been accounted for. In fact, without introducing any further approximation, we take the convolution of both sides of Eq. 4.66 with the $\phi_i(\vec{r};\vec{s})$ functions and we obtain the integrals required to setup and solve the PCM electrostatic problem,

$$\langle i|\mu\nu\rangle \simeq \frac{1}{2}S_{\mu\nu}\int d\vec{r}\,\phi_i(\vec{r};\vec{s}_i)\left(\frac{1}{|\vec{r}-\vec{R_{(\mu)}}|}+\frac{1}{|\vec{r}-\vec{R_{(\nu)}}|}\right). \tag{4.67}$$

Note that the integrals on the right hand side of Eq. 4.67 do not depend on the AO basis functions, they are the only kind of integrals specifically required by the combination of DFTB and PCM and they are easily evaluated by the same integral technology involved for *ab initio* calculations.

The interaction energy between $\tilde{n}(\vec{r})$ and $\sigma(\vec{s})$ in discretized form becomes

$$\tilde{E}_{int} = \sum_{Ai}\Delta q_A\gamma_{Ai}(|\vec{R}_A-\vec{s}_i|)q_i \tag{4.68}$$

which is just the classical electrostatic interaction between two distributions of point charges mediated by a isotropic Coulombic potential γ_{Ai}, which contains all the details of the physics

$$\gamma_{Ai}(|\vec{R}_A-\vec{s}_i|) = \langle A|i\rangle. \tag{4.69}$$

The notation γ_{Ai} has been chosen in analogy with the SCC-DFTB formalism, where the electrostatic Coulombic interaction between two atoms has been indicated as γ_{AB} (see Elstner et al.[111]). It is easy to recognize that the γ_{Ai} functions are just the integrals of the right hand side of Eq. 4.67, so that the evaluation of the DFTB potential $\tilde{V}_i$ at the solute-solvent interface can be recasted as

$$\tilde{V}_i = \frac{1}{2}\sum_{\mu\nu}P_{\mu\nu}S_{\mu\nu}\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)-\sum_A Z_A\gamma_{Ai} \tag{4.70}$$

while the PCM contribution to the Fock matrix as in Eq. 4.57 becomes

$$v_{\mu\nu}^{PCM}(\mathbf{P}) = \frac{1}{2}S_{\mu\nu}\sum_i\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)w_i(\mathbf{P}). \tag{4.71}$$

In Eq. 4.71 we made explicit the dependence of the polarization weights on $\mathbf{P}$ through the electrostatic potential (see Eq. 4.54), while it has been left implicit the dependence on the set of nuclear charges $\{Z_A\}$. So, each element $S_{\mu\nu}$ of the overlap matrix is multiplied by the potential generated by the solvent charges at (μ) and (ν) *nuclei* and the factor $\frac{1}{2}$ originates from the symmetry of $\mathbf{P}$ and $\mathbf{S}$ matrices. $v_{\mu\nu}^{PCM}$ is then a pure Coulombic contribution, and its form is very similar to that of the electrostatic interaction in the bielectronic term of the gas-phase Fock matrix (see Elstner et al.[111] and Zheng et al.[112])

$$G_{\mu\nu}^{0} = \frac{1}{2} S_{\mu\nu} \sum_{C} \left(\gamma_{(\mu)\,C} + \gamma_{(\nu)\,C} \right) q_C \tag{4.72}$$

as these two contributions to the Fock matrix, $v_{\mu\nu}^{PCM}$ and $G_{\mu\nu}^{0}$, only differ by the definition of the gamma functions $\gamma_{(\mu)C}$ and $\gamma_{(\mu)i}$. Eq. 4.68 and 4.71 are the basic ingredients for the implementation of the DFTB/PCM model.

In the last part of this section we add some further considerations about the electrostatics of the DFTB method, independently of its combination with the PCM model. In particular we present what we believe is a consistent way to evaluate the dipole and higher multipole moments within the DFTB scheme, without resorting to additional parameterization. The dipole moment associated to the charge density of Eq. 4.47 is

$$\vec{\mu} = \sum_{\mu\nu} P_{\mu\nu} \int d\vec{r}\, \chi_\mu(\vec{r})\, \vec{r}\, \chi_\nu(\vec{r}) - \sum_{A} Z_A \vec{R}_A \tag{4.73}$$

which requires the evaluation of the integrals

$$\int d\vec{r}\, \chi_\mu(\vec{r})\, \vec{r}\, \chi_\nu(\vec{r}), \tag{4.74}$$

which are not provided within the usual parameterization of DFTB. Therefore, we apply the previously described approach also to the calculation of the dipole moment, which can be approximated as

$$\tilde{\vec{\mu}} = \int d\vec{r}\; \vec{r}\; \tilde{n}(\vec{r}) = \sum_{A} \Delta q_A \vec{R}_A = \sum_{A} \sum_{\substack{\mu\nu \\ \mu\in A}} \frac{1}{2} (P_{\mu\nu} S_{\mu\nu} + P_{\nu\mu} S_{\nu\mu}) - \sum_{A} Z_A \vec{R}_A \tag{4.75}$$

obtaining the following final expression

$$\tilde{\vec{\mu}} = \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} S_{\mu\nu} \left(\vec{R}_{(\mu)} + \vec{R}_{(\nu)} \right) - \sum_{A} Z_A \vec{R_A}. \tag{4.76}$$

If we compare the dipole integrals of Eq. 4.73 with the approximate ones in Eq. 4.76, we find that the approximate dipole moment will be close to the *exact* one as long as

$$\int d\vec{r}\, \chi_\mu(\vec{r})\, \vec{r}\, \chi_\nu(\vec{r}) \simeq \frac{1}{2} S_{\mu\nu} (\vec{R_{(\mu)}} + \vec{R_{(\nu)}}). \tag{4.77}$$

If we evaluate these integrals approximating the atomic functions χ_μ and χ_ν with gaussian functions (see Appendix C), we find that Eq. 4.77 is always verified, without any restriction on the domain of $\vec{r}$. This suggests that the DFTB dipole moments can be computed using Eq. 4.76 rather than Eq. 4.73, and the accuracy should be comparable to that of a DFT calculation using a valence minimal basis set. All multipole moments up to sixth order have been implemented with this approach.

As a final note, we point out that no explicit PCM contribution needs to be added to Eq. 4.76 because the dipole moment is a first order property. In other words, when the energy in Eq. 4.58 (including the PCM contribution) has been variationally minimized, the dipole moment is simply computed from the resulting density matrix and the quantities on right hand side of Eq. 4.77.

4.2.3 Analytical derivatives of the DFTB/PCM energy

Using the approximations and the expressions introduced in the first part of this section, we now describe how to compute all the other quantities related to the electrostatic potential and needed for the evaluation of DFTB/PCM energy, its analytical first and second derivatives, and its derivative as require to solve the coupled perturbed SCF equations. The theory and implementation of the analytical derivatives of the PCM energy has been already reported in the literature[213,214,216,217] and here we will follow the formalism described by Barone et al.[213]

We already encountered the first example of derivative of the PCM energy, since the PCM contribution to the Fock matrix is the derivative of Eq. 4.68 with respect to the elements $P_{\mu\nu}$ of the density matrix. The discussion of all the relevant aspects has been reported in the previous section (see the discussion about Eq. 4.71), so we will not come back on this issue again.

In order to compute the analytical gradient of the energy of Eq. 4.68, we need to define the derivatives of the DFTB electrostatic potential ($\tilde{V}_i^x$) with respect to an external perturbation x, corresponding to the displacement of one of the nuclei. The derivative of gamma function γ_{Ai} involves two contributions: one arises from the perturbation of the atomic coordinate $\vec{R}_A$ and the other from the change in the location of the surface element $\vec{s}_i$, which ultimately depends on the position of the nuclei through the details of the discretization of the solute-solvent interface. More formally, we can write

$$(\gamma_{Ai})^x = \langle A^x|i\rangle + \frac{\partial\langle A|i\rangle}{\partial\vec{s}_i}\cdot\frac{\partial\vec{s}_i}{\partial x} = \langle A^x|i\rangle + \langle A|i'\rangle\cdot\frac{\partial\vec{s}_i}{\partial x}. \tag{4.78}$$

Thus, since any contribution from the density matrix derivative $\mathbf{P}^x$ vanishes for a variationally optimized wave-function, the final form of $\tilde{V}_i^x$ is

$$\tilde{V}_i^x = \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}S_{\mu\nu}^x\left(\gamma_{(\mu)\,i} + \gamma_{(\nu)\,i}\right) + \sum_A \Delta q_A\langle A^x|i\rangle + \sum_A \Delta q_A\langle A|i'\rangle\cdot\frac{\partial\vec{s}_i}{\partial x} \tag{4.79}$$

As already noted, all the integrals involved in the above equation, i.e. $\langle A^x|i\rangle$ and $\langle A|i'\rangle$, can be obtained using standard integral codes, while $S^x_{\mu\nu}$ can be computed from the derivatives of the analytical functions fitting the DFTB overlap parameters.

The evaluation of the density matrix derivative $\mathbf{P}^x$ becomes necessary in order to compute higher order molecular properties, such as harmonic vibrational frequencies, infrared intensities, vertical excitation energies within the linear response approach. The PCM contribution to the kernel of the CP-SCF equations corresponds to a Fock matrix contribution of the form $v^{PCM}_{\mu\nu}(\mathbf{P}^x)$ where the polarization weights are computed using the differential density matrix, i.e.

$$v^{PCM}_{\mu\nu}(\mathbf{P}^x) = \frac{1}{2}S_{\mu\nu}\sum_{ij}\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)\left(\frac{\mathbf{Q}+\mathbf{Q}^\dagger}{2}\right)_{ij}\tilde{V}_j(\mathbf{P}^x) \tag{4.80}$$

It is worth to note that $\tilde{V}_j(\mathbf{P}^x)$ is only the electronic component of the electrostatic potential, since the nuclear part does not depend on $\mathbf{P}^x$. When the perturbation is an external static electric field, the right hand side of the CP-SCF equations is given by the dipole integrals, which we compute as described in Eq. 4.76 for the DFTB method. The solution of such equations leads to the static polarizability tensor. On the other hand, when we include perturbations of the solute geometry, the right hand side of the CP-SCF equations involves the derivatives of the PCM contribution to the Fock matrix with respect to the nuclear displacements, which can be calculated from Eq. 4.71 and Eq. 4.78 together with other quantities such as the geometrical derivatives of the PCM matrix $\mathbf{Q}^x$,

$$\begin{aligned}
v^{PCM,\,x}_{\mu\nu}(\mathbf{P}) &= \frac{1}{2}S^x_{\mu\nu}\sum_{ij}\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)\left(\frac{\mathbf{Q}+\mathbf{Q}^\dagger}{2}\right)_{ij}\tilde{V}_j(\mathbf{P}) \\
&+ \frac{1}{2}S_{\mu\nu}\sum_{ij}\left((\gamma_{(\mu)\,i})^x+(\gamma_{(\nu)\,i})^x\right)\left(\frac{\mathbf{Q}+\mathbf{Q}^\dagger}{2}\right)_{ij}\tilde{V}_j(\mathbf{P}) \\
&+ \frac{1}{2}S_{\mu\nu}\sum_{ij}\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)\left(\frac{\mathbf{Q}^x+\mathbf{Q}^{x\,\dagger}}{2}\right)_{ij}\tilde{V}_j(\mathbf{P}) \\
&+ \frac{1}{2}S_{\mu\nu}\sum_{ij}\left(\gamma_{(\mu)\,i}+\gamma_{(\nu)\,i}\right)\left(\frac{\mathbf{Q}+\mathbf{Q}^\dagger}{2}\right)_{ij}\tilde{V}^x_j(\mathbf{P})
\end{aligned} \tag{4.81}$$

Finally, the remaining contributions to the Hessian of the energy are rather cumbersome to assemble, but the only new quantity that needs to be evaluated is the second derivative of γ_{Ai} which corresponds to

$$(\gamma_{Ai})^{x,y} = \langle A^{xy}|i\rangle + \langle A^x|i'\rangle\cdot\frac{\partial\vec{s}_i}{\partial y} + \langle A^y|i'\rangle\cdot\frac{\partial\vec{s}_i}{\partial x} + \frac{\partial\vec{s}_i}{\partial x}\cdot\langle A|i''\rangle\cdot\frac{\partial\vec{s}_i}{\partial y} + \langle A|i'\rangle\cdot\frac{\partial^2\vec{s}_i}{\partial x\partial y}. \tag{4.82}$$

As in the case of $(\gamma_{Ai})^x$ we stress that, all the additional integrals involved in the expression above, namely $S^{xy}_{\mu\nu}$, $\langle A^x|i'\rangle$ and $\langle A|i''\rangle$, are typically available from standard integral codes, while $S^{xy}_{\mu\nu}$ can be obtained from the DFTB parameters.

4.3 The TD-DFTB approach

The calculation of the excited state energies can be carried out using the Time-Dependent DFTB (TD-DFTB) method[134,198] which is based on the linear response theory applied to the DFTB ground state energy. Our recent implementation of the TD-DFTB[134] follows the formalism of Casida et al.[218,219] and, in contrast to a previous implementation,[198] does not require the inclusion of additional parameters with respect to the ground state theory. The form of the TD-DFTB equations is analogous to the corresponding one of the TD-DFT

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{4.83}$$

The eigenvalues Ω are the vertical excitation energies, while the amplitudes for the single particle excitation and de-excitation are contained in the eigenvectors $\mathbf{X}$ and $\mathbf{Y}$, respectively. The response matrices $\mathbf{A}$ and $\mathbf{B}$ are defined as

$$\begin{aligned} A_{ai\sigma,bj\sigma'} &= (\epsilon_{a\sigma} - \epsilon_{i\sigma})\delta_{ab}\delta_{ij}\delta_{\sigma\sigma'} + \frac{\partial F^{\sigma}_{ai}}{\partial P^{\sigma'}_{bj}} \\ B_{ai\sigma,bj\sigma'} &= \frac{\partial F^{\sigma}_{ai}}{\partial P^{\sigma'}_{jb}} \end{aligned} \tag{4.84}$$

where i, j are occupied orbitals, a, b are unoccupied ones and σ is a spin label. The usual procedure for the diagonalization of Eq. 4.83 involves the contraction of the combinations $(\mathbf{A}+\mathbf{B})$ and $(\mathbf{A}-\mathbf{B})$ with the $(\mathbf{X}+\mathbf{Y})$ and $(\mathbf{X}-\mathbf{Y})$ vectors, carried out in the AO basis.

Using the definitions given in Eq. 4.40 and Eq. 4.41 the Fock matrix derivatives with respect to the density matrix in the AO basis, can be written[134] in terms of a four-index Coulomb-like interaction kernel ($\mathbb{J}$) and a spin dependent exchange-like interaction kernel ($\mathbb{K}$)

$$\frac{\partial F^{\sigma}_{\mu\nu}}{\partial P^{\sigma'}_{\kappa\lambda}} = \mathbb{J}_{\mu\nu\kappa\lambda} + \mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} \tag{4.85}$$

being defined as

$$\mathbb{J}_{\mu\nu\kappa\lambda} = \frac{1}{4} S_{\mu\nu} \left(\gamma_{AC} + \gamma_{BC} + \gamma_{AD} + \gamma_{BD}\right) S_{\kappa\lambda} \tag{4.86}$$

$$\mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} = \frac{1}{4}\delta_{\sigma}\delta_{\sigma'} S_{\mu\nu} \left(W_{All''}\delta_{AC} + W_{Bl'l''}\delta_{BC} + W_{All'''}\delta_{AD} + W_{Bl'l'''}\delta_{BD}\right) S_{\kappa\lambda} \tag{4.87}$$

where the functions related to μ, ν, κ, and λ indexes are respectively located on A, B, C, and D atoms, and have l, l', l'', and l''' angular momenta. Also the total DFTB energy can be rewritten in terms of the interaction kernels[134]

$$E = \sum_{\sigma}\sum_{\mu\nu} h_{\mu\nu} P^{\sigma}_{\mu\nu} + \frac{1}{2}\sum_{\sigma\sigma'}\sum_{\mu\nu\kappa\lambda} \left[\mathbb{J}_{\mu\nu\kappa\lambda} + \mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda}\right] P^{\sigma}_{\mu\nu} P^{\sigma'}_{\kappa\lambda} + \frac{1}{2}\sum_{A\neq B} \left(V_{AB} + Z_A \gamma_{AB} Z_B\right). \tag{4.88}$$

Then, it is easy to realize that $(A+B)_{\mu\nu\sigma\kappa\lambda\sigma'}$ involves both charge and the symmetric component of spin interaction kernels,

$$(A+B)_{\mu\nu\sigma\kappa\lambda\sigma'} \leftarrow (\mathbb{J}_{\mu\nu\kappa\lambda} + \mathbb{J}_{\mu\nu\lambda\kappa}) + \left(\mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} + \mathbb{K}^{\sigma\sigma'}_{\mu\nu\lambda\kappa}\right) = 2\mathbb{J}_{\mu\nu\kappa\lambda} + \left(\mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} + \mathbb{K}^{\sigma\sigma'}_{\mu\nu\lambda\kappa}\right) \tag{4.89}$$

while $(A-B)_{\mu\nu\sigma\kappa\lambda\sigma'}$ involves only the antisymmetric component of the spin interaction kernel,

$$(A-B)_{\mu\nu\sigma\kappa\lambda\sigma'} \leftarrow (\mathbb{J}_{\mu\nu\kappa\lambda} - \mathbb{J}_{\mu\nu\lambda\kappa}) + \left(\mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} - \mathbb{K}^{\sigma\sigma'}_{\mu\nu\lambda\kappa}\right) = \left(\mathbb{K}^{\sigma\sigma'}_{\mu\nu\kappa\lambda} - \mathbb{K}^{\sigma\sigma'}_{\mu\nu\lambda\kappa}\right) \tag{4.90}$$

When the solvation effects are added to the linear response theory according the PCM scheme,[220] the four-index quantities in the above equations are supplemented by an explicit PCM contribution

$$\frac{\partial F^{\sigma}_{\mu\nu}}{\partial P^{\sigma'}_{\lambda\tau}} = \mathbb{J}_{\mu\nu\lambda\tau} + \mathbb{K}^{\sigma\sigma'}_{\mu\nu\lambda\tau} + \frac{\partial v^{PCM}_{\mu\nu}}{\partial P^{\sigma'}_{\lambda\tau}}. \tag{4.91}$$

which is a pure Coulomb-like term not depending on the spin label

$$\frac{\partial v^{PCM}_{\mu\nu}}{\partial P^{\sigma'}_{\lambda\tau}} = \frac{1}{4} S_{\mu\nu} S_{\lambda\tau} \sum_{ij} \left[\gamma_{(\mu)\,i} + \gamma_{(\nu)\,i}\right] \left(\frac{\mathbf{Q} + \mathbf{Q}^{\dagger}}{2}\right)_{ij} \left[\gamma_{(\lambda)\,j} + \gamma_{(\tau)\,j}\right] \tag{4.92}$$

The extensions of the TD-DFTB/PCM approach to the case of non-equilibrium solvation[98–100] is straightforward. The equilibrium $\mathbf{Q}(\varepsilon)$ matrix, which depends on the static dielectric constant ε, is simply replaced by $\mathbf{Q}(\varepsilon_\infty)$, where the optical dielectric constant ε_∞ is used in place of the static one.

Chapter 5

Environmental effects: results and discussion

In order to test the performances of the DFTB/PCM scheme, a benchmark of 18 neutral molecules and 4 ions has been chosen. DFTB method in its recently developed analytical formulation (DFTBA[112,131,134]) and other popular semiempirical methods (PM6[106] and AM1[105]), have been compared with selected DFT methods. In particular, we chose two hybrid functionals, B3LYP[66] and B2PLYP,[67,69] which are well known to be very accurate for the calculation of gas-phase molecular properties,[66,69,71,72] and the conventional PBE functional,[221] which has been shown to provide solvation energies in very good agreement with the experimental data.[213,222] Furthermore, the set of parameters employed by the DFTBA method has been originally fitted to the latter one. In conjunction with these functionals, we employed the 6-31G(d), 6-31G(d,p) and aug-cc-pVTZ (AVTZ) basis sets. In general, PCM calculations have been performed using Universal Force Field cavities with van der Waals atomic radii, scaled by a factor $\alpha = 1.1$. Different choices of cavity parameters will be explicitly indicated in the next sections. Anharmonic frequency calculations have been performed by second order perturbative (VPT2) vibrational approach.[23–25]

5.1 Energetics

The Solvation Energy (SE) for a given molecule can be defined, to a first approximation, as the difference between the total energy of the molecule optimized in vacuum and the free energy of solvation of the molecule optimized in water. Such an approach is formally exact, under the assumption that solvent effects do not significantly influence the rotational and vibrational degrees of freedom of the molecule. The free energy of solvation for DFT methods is calculated as in Eq. 4.46, using the electrostatic potential of Eq. 4.51. On the other side, DFTBA method employs the formulas discussed in the previous section for the energy (Eq. 4.58) and for the molecular potential (Eq. 4.70). Furthermore the DFTBA employs the minimal basis set

approximation, and the parameters of DFTBA have been fitted to PBE data. In order to separate the effects due to the different approximations in the DFTBA/PCM scheme, in Table 5.1 a set of values of SEs for a water molecule in aqueous solution have been reported, as computed with B3LYP/AVTZ, PBE/AVTZ, PBE/STO-3G and DFTBA methods. We used different scaling factors (α) in order to modify the cavity radii and computed the SE for nine values of α, ranging from 0.5 to 1.3. In Figure 5.1 the energies have been also plotted for a more clear visualization. The SEs computed at B3LYP/AVTZ and PBE/AVTZ levels are nearly identical for every value of α, suggesting that the effects due to the different exchange-correlation functionals should be very small. A more sensible difference has been found between the PBE/AVTZ and the PBE/STO-3G SEs. Such a difference is nearly constant when varying the cavity radii, and can be attributed to the minimal basis set approximation. The SEs computed at DFTBA level are very close to PBE/STO-3G for large cavities, then diverge for values of α lower than 0.7. This suggests that for large cavities, the energetics of the DFTBA method is mainly influenced by the minimal basis set approximation, while reducing the cavities the approximations introduced in the definition of the DFTBA molecular potential ($\tilde{V}$), discussed in the previous section, become important. In order to quantitatively compare molecular potentials computed with different methods, we estimated the Solvation Energies (SE0) by single point calculations at fixed geometry (i.e. a water molecule in gas-phase optimized at B3LYP/AVTZ level), without any further optimization. In Table 5.1 the SE0 have been reported as calculated by the mentioned methods. Now, for each α value and indipendently on the particular computational method employed, the vectors of the molecular potential at fixed geometry ($\mathbf{V^0}$) have the same number of elements and the coordinates of the surface elements are identical. The Mean Absolute Error (MAE) of PBE/AVTZ, PBE/STO-3G and DFTBA potentials with respect to B3LYP/AVTZ calculations, MAE($\mathbf{V^0}$), can then be computed, averaging the errors of each element of $\mathbf{V^0}$ over all the surface elements. The MAE($\mathbf{V^0}$) have been reported in Table 5.1 in atomic units. Furthermore, the values of the MAE($\mathbf{V^0}$) and the absolute values of the SE0s have been plotted together in Figure 5.2 for a visual comparison. From Table 5.2 it is evident that the approximations made in the PCM integrals (Eq. 4.67) become important only for very small values of the cavity radii, while in a wide range of distances the agreement between DFTBA and PBE/STO-3G is quantitative, both for the energies and the potential. Moreover, we should remember that for small cavities, the errors related to the molecular potential affect also the standard DFT methods, as long as the tails of the electronic densities are non-zero beyond the surface, where the integrals are evaluated.

In Table 5.2, the SEs for a set of 18 neutral molecules and 4 ions have been reported, as calculated at DFTBA level, with other semi-empirical methods (AM1 and PM6), PBE/STO-3G and other popular DFT methods. All the MAE reported in Table 5.2 have been computed with respect to B3LYP/AVTZ energies. In analogy with the previous discussion about the water molecule, we observe that the SEs at DFTBA level are comparable to those at PBE/STO-3G level, being the over-

all MAE of the DFTBA method for neutral molecules (1.47 kcal/mol), similar to PBE/STO-3G (1.24 kcal/mol). For a direct comparison between computed and experimental SEs, the effects due to electrostatic, cavitation and dispersion-repulsion interactions should be included[222] in the calculations. In the implementation of DFTB with COSMO model[223] (DFTB/COSMO), Elstner et al. fitted the atomic cavity parameters in order to reproduce the experimental SEs, obtaining a final rms error of 1.1 kcal/mol. Similar procedures have been used also in other solvation models developed for the DFTB method. In particular, the Poisson-Boltzmann model (DFTB/PB[224]) showed Mean Unsigned Errors with respect to experimental SE of 1.7 and 1.9 kcal/mol, for two different benchmark sets of molecules; finally, using the Generalized Born model (DFTB/GBSA[225]) Liu et al. obtained a final MAD of 1.11 kcal/mol. In both these two latter implementations, the atomic parameters defining the cavities and the surface tension parameters were fitted in order to reproduce the experimental energies. A fitting procedure similar to that employed by Elstner et al. could be in principle applied also to the DFTB/PCM model, in order to improve the energetics of solvation. Preliminary calculations showed that an energetics closer to DFT can be obtained by tuning the cavity radii. In particular, in Table 5.2 we note that using a reduced value of α (0.9 for Hydrogen, Carbon and Oxygen atoms, and of 0.8 for Nitrogen atoms), the accuracy of the SEs of the neutral molecules is improved, resulting in an overall MAE of just 0.67 kcal/mol with respect to B3LYP/AVTZ calculations. It is worth to note that the present empirical correction does not improve at the same time the SEs of both the neutral and ionic systems. More in general, the parameters defining the PCM cavities are very sensitive to the total charge of the solute, to the hybridization of the atoms and to substituent effects, and targeted studies are necessary in order to obtain an high accuracy in the energetics of solvation (see for example the discussion about the optimization of the cavity parameters for the UAHF model[215] and references therein).

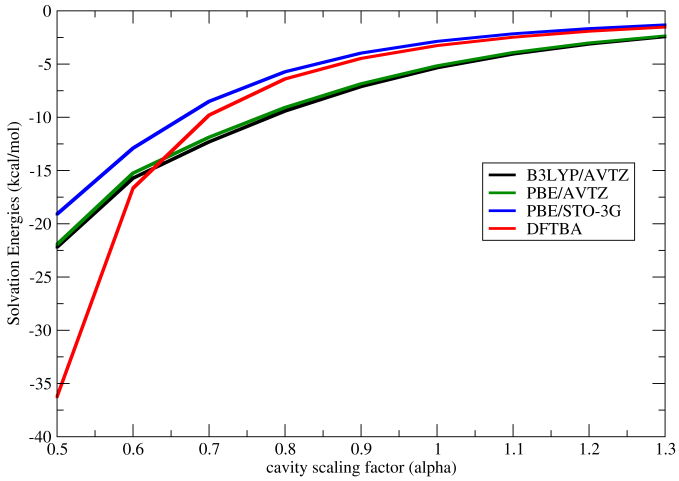


Figure 5.1: Solvation Energies (kcal/mol) of a water molecule in aqueous solution, computed with four different computational methods and plotted in function of the cavity scaling factor.

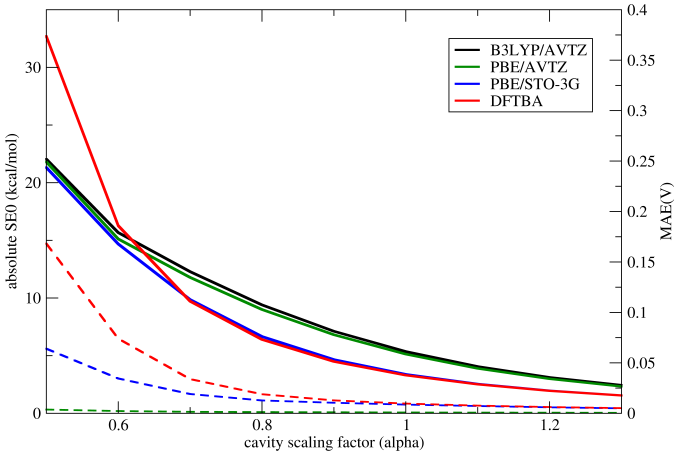


Figure 5.2: Straight lines refer to the absolute values of the Solvation Energies at fixed geometry (SE0) of a water molecule in aqueous solution (y axis on the left), as computed at B3LYP/AVTZ, PBE/AVTZ, PBE/STO-3G and DFTBA level. Corresponding values of MAE($\mathbf{V^0}$), computed with respect to B3LYP/AVTZ electrostatic potential, have been reported in dotted lines (y axis on the right). See Section 5.1 for details about the definitions of SE0 and MAE($\mathbf{V^0}$), and see Table 5.1 for numerical values.

	SE				SE0				MAE($\mathbf{V}^0$)		
α	B3LYP/AVTZ	PBE/AVTZ	PBE/STO-3G	DFTBA	B3LYP/AVTZ	PBE/AVTZ	PBE/STO-3G	DFTBA	PBE/AVTZ	PBE/STO-3G	DFTBA
0.5	-22.20	-21.94	-19.10	-36.25	-22.04	-21.80	-21.32	-32.70	0.0037	0.0640	0.1682
0.6	-15.72	-15.25	-12.90	-16.66	-15.67	-15.12	-14.68	-16.27	0.0022	0.0345	0.0740
0.7	-12.31	-11.88	-8.51	-9.80	-12.28	-11.78	-9.86	-9.72	0.0015	0.0191	0.0339
0.8	-9.42	-9.08	-5.73	-6.40	-9.40	-9.00	-6.66	-6.40	0.0012	0.0127	0.0189
0.9	-7.12	-6.86	-3.97	-4.47	-7.11	-6.81	-4.66	-4.49	0.0010	0.0104	0.0128
1.0	-5.36	-5.17	-2.87	-3.27	-5.36	-5.13	-3.38	-3.30	0.0008	0.0086	0.0096
1.1	-4.07	-3.92	-2.16	-2.48	-4.06	-3.89	-2.54	-2.51	0.0007	0.0072	0.0076
1.2	-3.13	-3.02	-1.67	-1.92	-3.12	-2.99	-1.96	-1.95	0.0006	0.0059	0.0062
1.3	-2.44	-2.36	-1.32	-1.53	-2.44	-2.33	-1.55	-1.55	0.0005	0.0049	0.0051

Table 5.1: Solvation Energies (SE) and Solvation Energies at fixed geometry (SE0) of a Water molecule (kcal/mol). SE0 calculated on gas-phase B3LYP/AVTZ geometry. α is the cavity scaling factor. MAE($\mathbf{V}^0$) is reported in atomic units (see Section 5.1 for the definition).

Solute	DFTBA		PBE/ STO-3G	PM6	AM1	PBE/ 6-31G(d)	B2PLYP/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31G(d,p)	B3LYP/ AVTZ
α	1.1	0.8/0.9	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Methane	-0.04	-0.12	-0.13	-0.17	-0.03	-0.14	-0.11	-0.11	-0.11	-0.12
Ethene	-0.24	-0.56	-0.37	-0.40	-0.27	-0.80	-0.81	-0.75	-0.72	-0.78
Benzene	-0.63	-1.34	-0.85	-1.35	-1.19	-1.80	-1.88	-1.66	-1.60	-1.55
Ammonia	-0.74	-1.91	-3.06	-4.44	-2.96	-3.86	-3.90	-3.82	-3.60	-3.08
Methylamine	-0.58	-1.57	-2.43	-3.94	-2.40	-2.85	-2.97	-2.86	-2.72	-2.61
Ethylamine	-0.55	-1.56	-2.40	-3.68	-2.44	-2.84	-3.00	-2.86	-2.71	-2.70
Piperidine	-0.42	-1.31	-1.93	-3.06	-1.96	-2.07	-2.27	-2.11	-2.03	-2.21
Pyridine	-1.04	-2.41	-1.96	-3.63	-2.67	-2.97	-3.33	-3.04	-3.01	-3.28
Pyrazine	-1.18	-2.94	-2.37	-4.71	-3.45	-3.15	-3.62	-3.32	-3.33	-3.69
Water	-2.48	-4.47	-2.16	-4.12	-3.14	-4.49	-4.83	-4.61	-4.39	-4.07
Methanol	-1.94	-3.62	-1.53	-3.56	-2.30	-2.98	-3.38	-3.15	-3.05	-3.13
Ethanol	-1.94	-3.76	-1.53	-3.76	-2.31	-2.85	-3.20	-2.99	-2.89	-3.07
Acetaldehyde	-2.39	-4.23	-1.04	-5.23	-3.57	-2.97	-3.77	-3.28	-3.30	-4.01
Acetone	-2.91	-5.09	-1.21	-5.42	-3.55	-3.38	-4.08	-3.65	-3.67	-4.45
Acetic Ac.	-3.47	-6.85	-1.84	-5.26	-4.01	-3.87	-4.61	-4.16	-4.14	-4.57
THF	-1.69	-3.18	-1.09	-3.48	-2.09	-1.92	-2.38	-2.22	-2.26	-2.43
Furan	-0.75	-1.61	-0.85	-1.48	-1.62	-1.90	-2.13	-1.90	-1.87	-1.89
Phenol	-1.91	-3.77	-2.36	-3.53	-2.81	-4.07	-4.33	-4.01	-3.88	-3.76
MAE(neutral)	1.47	0.67	1.24	0.68	0.48	0.38	0.26	0.28	0.24	-
NH_4^+	-73.23	-86.99	-72.07	-73.32	-73.32	-72.50	-72.72	-72.61	-72.66	-72.77
OH^-	-85.48	-105.95	-81.63	-87.65	-85.06	-85.18	-85.77	-85.33	-85.17	-80.14
H_3O^+	-77.07	-90.51	-78.38	-78.52	-79.10	-78.04	-78.08	-78.01	-77.76	-77.61
$NH_3^{+\cdot}$	-74.68	-90.65	-73.65	-75.11	-75.03	-74.28	-74.48	-74.38	-74.39	-74.50
MAE(ions)	1.63	17.27	0.95	2.39	1.87	1.49	1.54	1.47	1.35	-

Table 5.2: Solvation Energies for a set of organic molecules in aqueous solution (kcal/mol). $\alpha = 0.8/0.9$ indicates a value of 0.9 has been used for H,C,O, and a value of 0.8 for N.

5.2 Dipole moments

In Table 5.3 the total dipole moments in vacuum, as computed with different methods, are reported. It is known that the minimal basis set approximation usually affects the accuracy of the DFTB dipole moments when calculated from the Mulliken atomic charges.[118,130] This is confirmed in our computations, where the MAE of the dipole moments of DFTBA and PBE/STO-3G are very similar, being respectively 0.43 and 0.53 Debye. It has been shown that use of Lowdin charges, leads only to a marginal refinement,[130] and for an effective treatment of the electrostatics, more accurate methods, such as the CM3 proposed by Truhlar et al.[129,130] should be used. A new generation of DFTB methods is under development, based on the use of the CM3 charges in place of the Mulliken charges in the SCF, and on a third order expansion of the total energy on charge fluctuations, in order to provide an improved electrostatic description at DFTB level (see e.g. Spiegelman et al.,[126,128,226] York et al.,[227,228] Elstner et al.[229–231]). Hence, a detailed discussion about the accuracy of the dipole moments is far beyond the aim of this work. We just remark here that at variance from the CM3 method, our implementation (Eq. 4.76) does not require any additional parametrization, and leads to a MAE of the dipole moments in a very good agreement with the other implementations based on the Mulliken charges. In particular, Sattelmeyer et al.[118] obtained a MAE of $\sim$ 0.4 Debye, averaging the errors of the DFTB dipole moments with respect to MP2/cc-pVTZ calculations, over a set of 47 organic molecules; Truhlar et al.[130] found a value of 0.56 Debye calculating the errors with respect to the experimental dipole moments of a set of 219 molecules. The accuracy of our results is also in good agreement with the work of Liu et al.,[225] where the DFTB dipole moments have been compared to B3LYP/6-31G* calculations. On average, our calculations suggest that DFTBA dipoles have an accuracy comparable to PM6 calculations and slightly lower than AM1. This is in line with the results reported by Stewart,[106] who found an average error of 0.38 Debye for PM6 and 0.26 Debye for AM1 calculations, with respect to the experimental dipole moments. When molecules in aqueous solution are considered (see Table 5.4), the interaction between the solute electronic densities and the solvent charge distribution leads to larger dipole moments. For PM6, AM1 and DFT methods, the mean increment of the dipole moments due to the solvent is about 0.4 Debye. In the case of DFTBA with standard cavities, the dipole moments in water are underestimated by about 0.7 Debye with respect to B3LYP/AVTZ, resembling the behaviour of PBE/STO-3G. Also in this case, similarly to the SE, reduced values of the cavity radii lead to an improvement of results (the MAE of DFTBA calculations using $\alpha = 0.8/0.9$ is reduced to 0.60 Debye).

Solute	DFTBA	PBE/ STO-3G	PM6	AM1	PBE/ 6-31G(d)	B2PLYP/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31G(d,p)	B3LYP/ AVTZ
Ammonia	0.86	1.98	2.25	1.85	1.94	1.94	1.91	1.85	1.49
Methylamine	0.65	1.70	2.03	1.49	1.47	1.53	1.47	1.42	1.29
Ethylamine	0.67	1.71	2.05	1.55	1.44	1.51	1.44	1.39	1.27
Piperidine	0.53	1.49	1.85	1.37	1.17	1.28	1.18	1.15	1.18
Pyridine	1.15	2.00	2.35	1.97	2.15	2.28	2.19	2.18	2.26
Water	1.63	1.55	2.07	1.86	2.07	2.16	2.10	2.04	1.85
Methanol	1.47	1.30	2.05	1.62	1.63	1.81	1.69	1.66	1.66
Ethanol	1.36	1.24	1.91	1.55	1.50	1.67	1.56	1.53	1.58
Acetaldehyde	2.35	1.56	3.21	2.69	2.54	2.86	2.64	2.65	2.88
Acetone	2.65	1.71	3.51	2.92	2.73	3.02	2.82	2.83	3.08
AceticAc.	1.77	0.87	2.07	1.89	1.53	1.68	1.58	1.61	1.78
THF	1.68	1.27	2.28	1.92	1.49	1.91	1.79	1.66	1.65
Furan	0.04	0.18	0.34	0.49	0.55	0.75	0.63	0.63	0.67
Phenol	1.12	1.31	1.36	1.23	1.34	1.41	1.35	1.34	1.27
Average	1.28	1.42	2.10	1.74	1.68	1.84	1.74	1.71	1.71
MAE	0.43	0.53	0.43	0.17	0.18	0.16	0.15	0.12	-

Table 5.3: Total Dipole Moments (Debye) for a set of organic molecules in vacuum.

Solute	DFTBA		PBE/ STO-3G	PM6	AM1	PBE/ 6-31G(d)	B2PLYP/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31G(d,p)	B3LYP/ AVTZ
α	1.1	0.8/0.9	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Ammonia	0.90	0.97	2.05	2.42	1.93	2.19	2.18	2.16	2.09	1.90
Methylamine	0.72	0.78	1.81	2.28	1.67	1.76	1.83	1.76	1.71	1.73
Ethylamine	0.75	0.82	1.84	2.32	1.73	1.76	1.83	1.75	1.71	1.75
Piperidine	0.63	0.71	1.66	2.19	1.61	1.52	1.63	1.53	1.50	1.69
Pyridine	1.48	1.78	2.34	3.04	2.55	2.82	2.97	2.87	2.86	3.12
Water	1.73	1.81	1.64	2.22	1.98	2.29	2.38	2.31	2.26	2.16
Methanol	1.64	1.76	1.41	2.31	1.82	1.90	2.10	1.97	1.94	2.05
Ethanol	1.59	1.74	1.37	2.26	1.82	1.83	2.03	1.91	1.87	2.06
Acetaldehyde	2.88	3.24	1.75	4.13	3.34	3.15	3.52	3.26	3.27	3.71
Acetone	3.31	3.76	1.94	4.62	3.68	3.49	3.82	3.57	3.60	4.09
AceticAc.	2.27	2.66	1.01	2.61	2.27	1.93	2.10	1.99	2.02	2.34
THF	1.94	2.15	1.42	2.74	2.24	2.14	2.35	2.20	2.21	2.43
Furan	0.02	0.03	0.18	0.46	0.58	0.67	0.93	0.78	0.78	0.87
Phenol	1.45	1.63	1.63	1.76	1.60	1.76	1.81	1.75	1.73	1.71
Average	1.52	1.70	1.57	2.53	2.06	2.09	2.25	2.13	2.11	2.26
MAE	0.74	0.60	0.73	0.34	0.20	0.24	0.14	0.20	0.19	-

Table 5.4: Total Dipole Moments (Debye) for a set of organic molecules in aqueous solution. $\alpha = 0.8/0.9$ indicates a value of 0.9 has been used for H,C,O, and a value of 0.8 for N.

5.3 Harmonic vibrational frequencies

In Table 5.5 the MAEs of the harmonic frequencies of each molecule in vacuum, computed with respect to B3LYP/AVTZ calculations have been reported. We observe that DFTBA and AM1 are the most effective semi-empirical methods for the computation of harmonic frequencies, with a MAE of about 70 cm^{-1} in both cases, while the PM6 method shows a sensibly lower accuracy (MAE of 224 cm^{-1}). As widely disscussed in the Introduction, we recall that the computation of vibrational frequencies at DFTB level is a field of large interest, and special parametrizations[119–121,230] have been developed in last years, in order to fit the DFTB harmonic frequencies to the experimental fundamentals, at the price of a reduced accuracy of the energetics. Unfortunately, a successful case in which the energetics and the vibrational properties were accurately reproduced with the same set of parameters has not still been reported in literature, and at the moment the most used parameters are designed in order to reproduce only selected molecular properties.

Harmonic frequencies for the same set of molecules in aqueous solution are reported in Table 5.6. We note that the overall accuracy is very similar in solvent as in vacuum calculations for all the employed methods, and again DFTBA and AM1 approaches provide a better agreement with DFT than PM6. Nevertheless, solvation effects are relevant for just few specific modes, being negligible for most of the other vibrations. For this reason, an analysis of the frequency shifts of such modes, is more meaningful than the analysis of the overall MAE.

In Table 5.7 the solvatochromic shifts of the C=O, N-H and O-H stretching modes in aqueous solution are reported as calculated with different methods. We note that for these modes solvent effects are significant, with shifts at B3LYP/AVTZ level of about 46 cm^{-1} for C=O stretching modes, 12 cm^{-1} for N-H stretching modes, and 19 cm^{-1} for O-H stretching modes. Among the semi-empirical methods, DFTBA with standard cavities is the most accurate method in the reproduction of such shifts. In particular, the solvent shifts of the C=O stretching modes are reproduced with errors lower than 5 cm^{-1} for Acetaldehyde, Acetic Acid and Acetone molecules. This is the most accurate result among either the other semi-empirical methods, and also when compared with the PBE/6-31G(d) method. Analogously, the average errors affecting the shifts of the N-H and O-H stretching modes are smaller than 9 and 7 cm^{-1} respectively. The maximum error of DFTBA calculations is related to the solvatochromic shift of the O-H stretching mode of Phenol, and it is just 16 cm^{-1}, much lower than the maximum errors of PM6 (32 cm^{-1}) and PBE/6-31G(d) (20 cm^{-1}) methods, and comparable with the maximum error of the AM1 method (17 cm^{-1}). In this respect we recall that, for aromatic molecules and conjugated systems, effects due to the lack of diffuse functions, self-interaction error and overpolarization are known to affect the DFTB treatment.[125,232–234]

As a final remark about harmonic frequency calculations, we point out that the calculations with standard cavities seem to be more reliable than the calculations with reduced cavities for vibrational properties. In particular we note that the

average errors on the solvatochromic shifts of C=O stretching modes are 3 cm^{-1} for $\alpha = 1.1$, and 38 cm^{-1} for $\alpha = 0.8/0.9$. Another analogous case is the shift of the O-H stretching modes of water molecule, which are sensibly better using $\alpha = 1.1$ than $\alpha = 0.8/0.9$. This suggests that a fitting procedure of the cavity parameters, similar to those employed in the DFTB/COSMO,[223] DFTB/PB[224] and DFTB/GBSA[225] implementations, could lead to an improved enegetics of solvation, at the price of a reduced accuracy for the vibrational properties. This is true in particular for the schemes which employ a low number of parameters, as our present case. So in our approach we prefer to avoid any fitting procedure, in order to have more flexibility in the calculations of the spectroscopical properties.

	DFTBA	PM6	AM1	PBE/6-31G(d)
acetaldehyde	76	152	71	28
aceticac	60	195	78	30
acetone	55	152	56	25
ammonia	42	374	83	78
benzene	74	114	41	33
ethanol	69	213	72	31
ethene	107	187	47	37
ethylamine	61	191	60	31
furan	81	128	64	33
methane	48	253	70	16
methanol	86	257	83	44
methylamine	73	234	67	37
phenol	62	123	66	35
piperidine	68	139	49	27
pyrazine	103	110	46	36
pyridine	85	109	43	35
thf	76	150	62	30
water	91	948	289	125
MAE	73	224	75	39

Table 5.5: Mean Absolute Errors (cm^{-1}) of harmonic frequencies of a set of organic molecules in vacuum, with respect to B3LYP/AVTZ.

	DFTBA		PM6	AM1	PBE/6-31G(d)
α	1.1	0.8/0.9	1.1	1.1	1.1
acetaldehyde	75	75	162	67	29
aceticac	58	61	205	80	30
acetone	53	56	161	52	26
ammonia	44	46	392	105	80
benzene	74	73	116	40	32
ethanol	67	67	203	73	31
ethene	107	108	189	47	37
ethylamine	59	58	194	61	29
furan	79	79	131	61	32
methane	41	42	260	69	17
methanol	86	88	267	83	43
methylamine	70	70	239	71	35
phenol	62	61	125	66	34
piperidine	67	66	140	49	27
pyrazine	103	99	116	48	36
pyridine	85	82	113	41	35
thf	75	72	157	65	29
water	91	97	932	287	116
MAE	72	72	228	76	39

Table 5.6: Mean Absolute Errors (cm^{-1}) of harmonic frequencies of a set of organic molecules in aqueous solution, with respect to B3LYP/AVTZ. $\alpha = 0.8/0.9$ indicates a value of 0.9 has been used for H,C,O, and a value of 0.8 for N.

		Solvent Shifts						Absolute Errors				
Mode		DFTBA		PM6	AM1	PBE/6-31G(d)	B3LYP/AVTZ	DFTBA		PM6	AM1	PBE/6-31G(d)
Mode	α	1.1	0.8/0.9	1.1	1.1	1.1	1.1	1.1	0.8/0.9	1.1	1.1	1.1
C=O	acetaldehyde	-43	-76	-56	-30	-28	-41	2	35	15	11	13
	aceticac	-57	-100	-65	-37	-35	-52	5	48	13	15	17
	acetone	-42	-76	-69	-31	-30	-45	3	31	24	14	15
	Average	-47	-84	-63	-32	-31	-46	3	38	17	13	15
N-H	ammonia	-6	-13	-7	-6	-11	-14	8	1	7	8	3
	”	-6	-14	-25	-17	-11	-14	8	0	11	3	3
	”	-5	-12	-25	-18	1	-5	0	7	20	13	6
	methylamine	-2	-6	-11	-9	-9	-15	13	9	4	6	6
	”	-2	-6	-21	-16	-3	-9	7	3	12	7	6
	ethylamine	-2	-5	-13	-5	-9	-15	13	10	2	10	6
	”	-2	-6	-18	-11	-3	-9	7	3	9	2	6
	piperidine	1	2	-14	-5	-5	-12	13	14	2	7	7
	Average	-3	-8	-17	-11	-6	-12	9	6	8	7	5
O-H	water	-23	-40	-1	-11	-13	-28	5	12	27	17	15
	”	-18	-34	-46	-22	6	-14	4	20	32	8	20
	methanol	-10	-21	0	-15	0	-15	5	6	15	0	15
	ethanol	-10	-20	0	-17	-1	-16	6	4	16	1	15
	phenol	-8	-18	-10	-31	-10	-24	16	6	14	7	14
	Average	-14	-27	-12	-19	-6	-19	7	10	21	7	16

Table 5.7: Harmonic frequencies and solvatochromic shifts (cm^{-1}) for a set of selected modes of organic molecules in aqueous solution. Absolute errors have been computed with respect to B3LYP/AVTZ.

5.4 Anharmonic vibrational frequencies

In the VPT2 framework, it has been shown that the anharmonic corrections are usually similar when computed with different computational approaches,[25,69,144] and hybrid schemes coupling different DFT methods or Coupled Cluster (CC) and DFT, have been found to provide very accurate vibrational properties when compared to experimental data, at reasonable computational cost.[6,71,235] In particular, for Uracil molecule[6] a CC/DFT hybrid scheme - where the anharmonic shifts computed at DFT level were combined with the CC harmonic frequencies - led to frequencies in very good agreement with experimental results. In analogy with our previous works, we can try to extend such an hybrid scheme from the combination of CC/DFT to the combination of DFT/DFTB methods.

In Table 5.8 and 5.9, the anharmonic shifts computed with different methods have been combined to the B3LYP/AVTZ harmonic frequencies, for Acetaldehyde and Acetic Acid molecules respectively. For each of these molecules, we note a good overall agreement between hybrid DFT/DFTBA frequencies and experimental fundamentals, with an overall MAE of 19 cm^{-1} for Acetaldehyde and 10 cm^{-1} for Acetic Acid. In the latter case, the performances of the hybrid DFT/DFTBA approach are even better than the DFT/PBE/6-31G(d) approach, which shows a MAE of 16 cm^{-1}. For comparison, an overall MAE of about 10 cm^{-1} was found when the hybrid CC/DFT scheme was applied to the calculation of the frequencies of Uracil molecule,[6] while calculations at full DFT level for gas-phase Glycine molecule[71] led to a MAE of 11 cm^{-1}. Furthermore, a MAE of 38 cm^{-1} was found when an approximate VPT2 scheme[50] was used in conjuction with full DFT calculations, in order to compute the infrared spectrum of the Glycine molecule adsorbed on the (100) surface of Silicon.[51] The results obtained at the DFTBA level are even more encouraging if we consider that the MAE can be mainly attributed to large errors localized on a few stretching modes, while all the other vibrations are well reproduced. In particular, the main sources of error for Acetaldehyde are two CH stretching modes, whose frequencies differ from the experimental values by 98 and 48 cm^{-1} respectively, and the C=O stretching mode (error of 47 cm^{-1}). The situation is even better for Acetic Acid molecule, for which a significant error (49 cm^{-1}) is found only for one CH stretching mode. Generally speaking, the discrepancies found in DFTBA calculations related to the frequencies of stretching modes of polarized bonds, could be attributed to the fair DFTBA electrostatic description of the chemical bonding, based on the Mulliken charges,[130] and on the minimal basis approximation. On the other side, the large errors of the CH stretching modes are very similar to those found by Witek et al.,[117,119] and they can be related to the parameters of the C and H interactions. So we suggest that a reparametrization of the repulsive potentials based on high level harmonic frequencies, rather than on experiment, could sensibly improve the hybrid frequencies for medium sized systems and could also lead to acceptable anharmonic frequencies at full DFTBA level, which could be used for very large systems. Finally we remark that one of the main advan-

tages of our VPT2 approach, compared to other time-independent approaches,[125,126] is that Fermi resonances can be effectively treated by the new HDCPT2 model.[25]

Table 5.8 – *Acetaldehyde*

				harmonic frequencies				
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	3027 (110)	2789 (348)	3154 (16)	3106 (31)	3134 (3)	3138	CH_3 stretching
11	A"	2996 (76)	2691 (380)	3089 (16)	3041 (31)	3072 (0)	3072	CH_3 stretching
2	A'	2871 (185)	2691 (365)	3062 (6)	2978 (78)	3021 (34)	3056	CH_3 stretching
3	A'	2472 (369)	2645 (196)	3057 (216)	2794 (47)	2868 (27)	2841	CH stretching
4	A'	1740 (33)	1795 (20)	2059 (285)	1787 (13)	1805 (30)	1774	C=O stretching
12	A"	1424 (45)	1315 (153)	1426 (43)	1449 (20)	1469 (0)	1469	CH_3 bending
5	A'	1422 (34)	1279 (177)	1380 (76)	1440 (16)	1459 (2)	1457	CH_3 bending
6	A'	1405 (5)	1228 (182)	1372 (38)	1398 (12)	1421 (10)	1411	CH bending
7	A'	1353 (46)	1211 (188)	1358 (41)	1346 (54)	1379 (20)	1400	CH_3 flip
13	A"	1124 (20)	1145 (0)	1218 (72)	1102 (43)	1135 (9)	1145	CH oop bending
8	A'	1062 (67)	1082 (47)	1118 (11)	1099 (30)	1129 (0)	1129	
9	A'	950 (43)	984 (78)	1033 (126)	870 (36)	886 (20)	906	CC stretching
14	A"	743 (37)	804 (23)	814 (33)	747 (33)	774 (5)	780	
10	A'	515 (5)	489 (31)	521 (0)	491 (29)	509 (11)	521	
15	A"	121	70	93	153	157	-	
MAE		77	156	70	34	12	0	
				anharmonic frequencies				
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	2923 (90)	2762 (251)	3058 (43)	2954 (60)	2991 (22)	3014	CH_3 stretching
11	A"	2904 (59)	2649 (314)	2976 (11)	2894 (69)	2933 (31)	2964	CH_3 stretching
2	A'	2820 (102)	2650 (272)	2965 (41)	2870 (52)	2917 (5)	2923	CH_3 stretching
3	A'	2417 (297)	2616 (99)	2963 (248)	2591 (123)	2673 (42)	2715	CH stretching
4	A'	1728 (17)	1776 (30)	2033 (287)	1762 (16)	1780 (34)	1746	C=O stretching
12	A"	1386 (50)	1298 (137)	1396 (39)	1407 (28)	1426 (9)	1436	CH_3 bending
5	A'	1387 (45)	1264 (168)	1350 (82)	1400 (32)	1421 (12)	1433	CH_3 bending
6	A'	1379 (15)	1200 (194)	1343 (51)	1370 (24)	1396 (1)	1394	CH bending
7	A'	1326 (25)	1191 (161)	1335 (17)	1313 (39)	1345 (6)	1352	CH_3 flip
13	A"	1103 (10)	1122 (8)	1191 (77)	1077 (36)	1111 (2)	1113	CH oop bending
8	A'	1042 (55)	1076 (21)	1107 (10)	1075 (22)	1104 (6)	1097	
9	A'	946 (80)	986 (120)	1026 (160)	831 (34)	852 (13)	865	CC stretching
14	A"	734 (29)	797 (33)	805 (40)	734 (29)	762 (1)	764	
10	A'	516 (7)	489 (19)	521 (12)	491 (17)	511 (2)	508	
15	A"	108	52	82	140	143	144	
MAE		63	131	80	41	13	0	

Continues on next page.

Table 5.8 – *Continues from previous page.*

		anharmonic shifts						
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	-104	-27	-96	-152	-142	-123	CH_3 stretching
11	A"	-92	-42	-113	-147	-139	-108	CH_3 stretching
2	A'	-50	-40	-97	-108	-104	-133	CH_3 stretching
3	A'	-55	-29	-94	-202	-195	-126	CH stretching
4	A'	-11	-18	-26	-25	-24	-28	C=O stretching
12	A"	-38	-17	-29	-41	-42	-33	CH_3 bending
5	A'	-34	-14	-29	-39	-38	-23	CH_3 bending
6	A'	-26	-28	-29	-27	-25	-16	CH bending
7	A'	-26	-20	-22	-32	-33	-47	CH_3 flip
13	A"	-21	-22	-27	-25	-24	-31	CH oop bending
8	A'	-20	-6	-10	-24	-24	-32	
9	A'	-3	1	-6	-38	-34	-40	CC stretching
14	A"	-8	-7	-9	-13	-12	-16	
10	A'	0	0	0	0	1	-12	
15	A"	-12	-18	-10	-13	-13	-	
		hybrid frequencies						
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	3030 (15)	3106 (92)	3037 (23)	2981 (32)	2991 (22)	3014	CH_3 stretching
11	A"	2980 (16)	3030 (65)	2959 (4)	2925 (38)	2933 (31)	2964	CH_3 stretching
2	A'	2971 (48)	2980 (57)	2924 (0)	2913 (9)	2917 (5)	2923	CH_3 stretching
3	A'	2813 (98)	2839 (124)	2774 (58)	2666 (49)	2673 (42)	2715	CH stretching
4	A'	1793 (47)	1786 (40)	1778 (32)	1779 (33)	1780 (34)	1746	C=O stretching
12	A"	1431 (5)	1451 (15)	1439 (3)	1427 (8)	1426 (9)	1436	CH_3 bending
5	A'	1425 (8)	1445 (12)	1430 (3)	1420 (12)	1421 (12)	1433	CH_3 bending
6	A'	1395 (0)	1393 (1)	1392 (2)	1394 (0)	1396 (1)	1394	CH bending
7	A'	1352 (0)	1358 (6)	1356 (3)	1346 (5)	1345 (6)	1352	CH_3 flip
13	A"	1114 (0)	1113 (0)	1108 (5)	1110 (3)	1111 (2)	1113	CH oop bending
8	A'	1109 (11)	1122 (25)	1119 (21)	1105 (7)	1104 (6)	1097	
9	A'	882 (16)	887 (21)	879 (13)	847 (18)	852 (13)	865	CC stretching
14	A"	766 (2)	767 (3)	765 (1)	761 (2)	762 (1)	764	
10	A'	509 (1)	509 (1)	509 (0)	510 (1)	511 (2)	508	
15	A"	144	139	146	144	143	144	
MAE		19	33	12	16	13	0	

Table 5.8: Harmonic, anharmonic (GVPT2), hybrid frequencies and anharmonic shifts in cm^{-1} of Acetaldehyde. In parenthesis the absolute errors with respect to experimental frequencies. Mean Absolute Errors (MAE) for harmonic and anharmonic frequencies have been computed with reference to experimental frequencies. Mode 15 has been neglected in the MAE calculation due to the lack of experimental harmonic frequency. Hybrid frequencies have been calculated by summing the harmonic frequencies at B3LYP/AVTZ level to the anharmonic shifts of each method. [a] Experimental harmonic and anharmonic frequencies from measurement on gas-phase Acetaldehyde performed by Wiberg et al.[236]

Table 5.9 – *Acetic Acid*

				harmonic frequencies				
		DFTBA	PM6	AM1	PBE/	B3LYP/		
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	3621 (116)	2780 (957)	3429 (307)	3561 (176)	3737		OH stretching
2	A'	3037 (122)	2682 (476)	3152 (7)	3129 (29)	3159		CH_3 stretching
13	A"	3003 (104)	2676 (431)	3068 (39)	3076 (31)	3108		CH_3 stretching
3	A'	2879 (173)	2550 (502)	3057 (5)	3009 (43)	3052		CH_3 stretching
4	A'	1703 (108)	1839 (27)	2087 (275)	1802 (9)	1811		C=O stretching
14	A"	1425 (52)	1319 (158)	1549 (71)	1456 (21)	1477		CH_3 bending
5	A'	1422 (49)	1303 (168)	1429 (42)	1452 (18)	1471		CH_3 bending
6	A'	1388 (19)	1217 (191)	1412 (4)	1381 (27)	1408		CH_3 flip
7	A'	1275 (59)	1203 (131)	1371 (36)	1320 (14)	1334		CO stretching
8	A'	1182 (18)	1147 (53)	1363 (162)	1182 (18)	1201		OH ip bending
15	A"	1027 (42)	1035 (35)	1100 (29)	1037 (33)	1070		
9	A'	991 (7)	1020 (22)	1072 (74)	968 (29)	998		
10	A'	866 (9)	951 (94)	1038 (180)	847 (10)	857		CC stretching
16	A"	588 (74)	593 (70)	589 (74)	682 (19)	663		OH oop bending
11	A'	560 (24)	513 (70)	569 (14)	561 (22)	584		
17	A"	483 (62)	504 (41)	520 (25)	530 (15)	546		
12	A'	401 (22)	375 (48)	418 (5)	406 (17)	424		
18	A"	60 (9)	48 (22)	24 (45)	61 (8)	70		
MAE		59	194	77	30	0		
				anharmonic frequencies				
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	3471 (96)	2752 (815)	3337 (230)	3372 (195)	3554 (13)	3568	OH stretching
2	A'	2931 (125)	2643 (413)	3055 (1)	2974 (82)	3015 (41)	3057	CH_3 stretching
13	A"	2908 (98)	2635 (371)	2970 (36)	2920 (86)	2966 (40)	3007	CH_3 stretching
3	A'	2829 (123)	2517 (435)	2964 (11)	2900 (52)	2944 (8)	2953	CH_3 stretching
4	A'	1673 (106)	1831 (51)	2038 (257)	1768 (12)	1779 (0)	1780	C=O stretching
14	A"	1394 (46)	1290 (150)	1513 (72)	1410 (30)	1438 (2)	1441	CH_3 bending
5	A'	1380 (55)	1285 (150)	1402 (33)	1415 (20)	1434 (1)	1436	CH_3 bending
6	A'	1358 (22)	1192 (188)	1385 (4)	1349 (31)	1369 (11)	1381	CH_3 flip
7	A'	1251 (73)	1183 (141)	1338 (13)	1261 (63)	1319 (5)	1325	CO stretching
8	A'	1153 (3)	1083 (66)	1333 (183)	1145 (4)	1155 (5)	1150	OH ip bending
15	A"	1007 (39)	1023 (23)	1093 (46)	1023 (23)	1052 (5)	1047	
9	A'	976 (9)	988 (2)	1051 (65)	953 (32)	980 (5)	986	
10	A'	844 (0)	940 (96)	1024 (180)	829 (14)	839 (4)	844	CC stretching
16	A"	564 (70)	575 (59)	564 (70)	659 (24)	642 (7)	635	OH oop bending
11	A'	555 (22)	513 (64)	571 (6)	553 (24)	576 (1)	577	
17	A"	469 (63)	499 (33)	508 (24)	520 (12)	531 (1)	533	
12	A'	401 (21)	374 (48)	417 (5)	407 (15)	424 (1)	423	
18	A"	52 (50)	31 (71)	-6 (109)	49 (53)	55 (47)	103	
MAE		57	176	75	43	11	0	

Continues on next page.

Table 5.9 – *Continues from previous page.*

		anharmonic shifts						
		DFTBA	PM6	AM1	PBE/	B3LYP/		
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	-149	-27	-92	-189	-182		OH stretching
2	A'	-105	-39	-96	-155	-144		CH_3 stretching
13	A"	-95	-41	-98	-156	-142		CH_3 stretching
3	A'	-50	-32	-93	-108	-107		CH_3 stretching
4	A'	-29	-7	-49	-34	-32		C=O stretching
14	A"	-30	-28	-35	-46	-39		CH_3 bending
5	A'	-42	-18	-27	-37	-37		CH_3 bending
6	A'	-30	-24	-27	-31	-38		CH_3 flip
7	A'	-23	-20	-33	-58	-15		CO stretching
8	A'	-29	-63	-30	-37	-45		OH ip bending
15	A"	-19	-11	-6	-14	-17		
9	A'	-14	-32	-21	-14	-17		
10	A'	-21	-11	-14	-17	-18		CC stretching
16	A"	-24	-17	-24	-22	-20		OH oop bending
11	A'	-4	0	1	-8	-8		
17	A"	-14	-4	-11	-10	-14		
12	A'	0	0	0	0	0		
18	A"	-7	-16	-31	-11	-15		
		hybrid frequencies						
		DFTBA	PM6	AM1	PBE/	B3LYP/	exp.[a]	
modes	symm.				6-31G(d)	AVTZ		assignments
1	A'	3588 (20)	3710 (142)	3645 (77)	3548 (19)	3554 (13)	3568	OH stretching
2	A'	3053 (3)	3120 (63)	3062 (5)	3004 (52)	3015 (41)	3057	CH_3 stretching
13	A"	3012 (5)	3067 (60)	3010 (3)	2952 (54)	2966 (40)	3007	CH_3 stretching
3	A'	3002 (49)	3019 (66)	2959 (6)	2943 (9)	2944 (8)	2953	CH_3 stretching
4	A'	1782 (1)	1804 (23)	1762 (18)	1777 (3)	1779 (0)	1780	C=O stretching
14	A"	1447 (6)	1449 (8)	1442 (1)	1431 (9)	1438 (2)	1441	CH_3 bending
5	A'	1429 (6)	1453 (17)	1444 (8)	1434 (1)	1434 (1)	1436	CH_3 bending
6	A'	1378 (2)	1383 (2)	1380 (0)	1377 (3)	1369 (11)	1381	CH_3 flip
7	A'	1310 (14)	1314 (10)	1301 (23)	1276 (48)	1319 (5)	1325	CO stretching
8	A'	1171 (21)	1137 (12)	1170 (20)	1163 (13)	1155 (5)	1150	OH ip bending
15	A"	1050 (3)	1058 (11)	1063 (16)	1056 (9)	1052 (5)	1047	
9	A'	983 (2)	966 (19)	977 (8)	983 (2)	980 (5)	986	
10	A'	835 (8)	845 (1)	843 (0)	839 (4)	839 (4)	844	CC stretching
16	A"	639 (4)	645 (10)	638 (3)	640 (5)	642 (7)	635	OH oop bending
11	A'	579 (1)	583 (6)	585 (8)	575 (2)	576 (1)	577	
17	A"	531 (1)	541 (8)	534 (0)	535 (2)	531 (1)	533	
12	A'	424 (1)	423 (0)	423 (0)	424 (1)	424 (1)	423	
18	A"	62 (40)	53 (49)	39 (63)	58 (44)	55 (47)	103	
MAE		10	28	14	15	11		

Table 5.9: Harmonic, anharmonic (GVPT2), hybrid frequencies and anharmonic shifts (cm^{-1}) of Acetic Acid. In parenthesis the absolute errors with respect to experimental frequencies. Mean Absolute Errors (MAE) for anharmonic frequencies have been computed with reference to experimental frequencies. Hybrid frequencies have been calculated by summing the harmonic frequencies at B3LYP/AVTZ level to the anharmonic shifts of each method. [a] Pettersson et al.[237]

5.5 Test case: vibrational frequencies of Uracil in aqueous solution

In Figure 5.3 the first shell of solvation of Uracil in aqueous solution is shown. This model, embedded into the polarizable continuum, has been widely used in literature (see Improta et al.[91] and references therein) and it has been adopted here in order to model the spectroscopic properties of solvated Uracil.

In Table 5.10 the harmonic frequencies and solvatocrhomic shifts of selected stretching modes of Uracil, as calculated at DFTBA and B3LYP/AN07D levels employing different solvent models, are reported. The shifts have been also visualized in Figure1 5.4 by histograms, for a better comparison of results. In this case we chose B3LYP/AN07D method as reference method, because it was found to provide very good results in the computation of vibrational frequencies of the gas-phase Uracil molecule.[6] In Table 5.10 we note that the absolute magnitude of the errors of the DFTBA harmonic frequencies is quite large, in line with the results of Table 5.5 and 5.6. However, it is noteworthy that the solvatochromic shifts of the frequencies calculated at DFTBA and B3LYP/AN07D are in very good qualitative agreement. In particular, when four explicit water molecules are added to the gas-phase Uracil molecule Figure 5.4, part b), both methods predict that the frequencies of the N-H and C=O stretching modes are redshifted, while the C-H stretching modes do not undergo any significant shift. When PCM is added to the gas-phase Uracil and Uracil+4H_2O Figure 5.4, part c), the agreement between the DFTBA and B3LYP/AN07D shifts is again reasonable, the shifts being always in the same direction. There is only one case in which the frequency shifts due to the PCM have different signs between DFT and DFTBA calculations: the N-H stretching modes of Uracil+4H_2O+PCM model. This could be due to the fact that the N-H moieties of Uracil are involved in the hydrogen bonds with the solvent molecules (see Figure 5.3), so that polarization effects induced by such molecules, not well described at the DFTBA level, could occur. However, since the PCM contribution to the total solvatochromic shift is very small when compared to the shift due to the inclusion of the explicit water molecules, the total shift is still in a reasonable agreement between DFT and DFTBA calculations, as shown in Figure 5.4, part a. The overall agreement between DFT and DFTBA solvent shifts suggests that the approach used to compute the hybrid DFT/DFTBA frequencies for gas-phase Acetaldehye and Acetic Acid molecules, could be extended to the computation of accurate hybrid frequencies in solution.

In Table 5.11 the hybrid frequencies of Uracil molecule, obtained adding the anharmonic shifts at DFTBA level to the B3LYP/AN07D harmonic frequencies, have been reported. For the gas-phase Uracil, we observe that the overall MAE, averaged over all normal modes, between hybrid and experimental frequencies is just 19 cm^{-1}, comparable to that observed for Acetaldehyde. The experimental spectra of Uracil in aqueous solution are available only in the region below 1800 cm^{-1}, due to the

strong absorptions of the O-H stretching modes of the solvent water molecules. So we selected nine normal modes, of frequencies between 1800 and 1000 cm^{-1}, and we used the available experimental frequencies for the study of the solvatochromic shifts computed at DFT/DFTBA level, using different solvation models. If we compare the MAE of Uracil+4H_2O (20 cm^{-1}) with the MAE of Uracil+PCM (36 cm^{-1}), we note that the inclusion of the four explicit water molecules sensibly improves the agreement with the experimental data. In particular, the C=O stretching modes and the mode 9 (in plane ring distortion) are sensibly improved by the explicit inclusion of water molecules. The MAE of Uracil+4H_2O+PCM is 25 cm^{-1}, comparable to Uracil+4H_2O. The PCM sensibly improves modes 10,12,13, which are quite delocalized over the Uracil molecule. These results seem to indicate that inclusion of some explicit water molecules is necessary in order to correctly compute the frequencies of localized modes, which involve a direct interaction with the potential of the water molecules. The PCM provides a general improvement of the low frequency, delocalized modes.

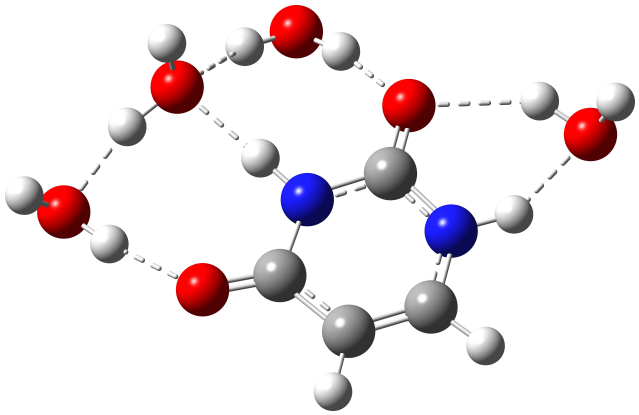


Figure 5.3: Four explicit water molecules of the first shell of solvation of Uracil molecule in aqueous solution.

stretching modes		gas-phase	Uracil +PCM	Uracil +4H_2O	Uracil +4H_2O +PCM
1. N-H	DFTBA(0.8/0.9)	3565	3561 (-4)	3233 [-332]	3290 (+57)
	DFTBA(1.1)	”	3558 (-7)	”	3203 (-30)
	B3LYP/AN07D	3639	3615 (-24)	3398 [-241]	3395 (-3)
2. N-H	DFTBA(0.8/0.9)	3545	3542 (-3)	3062 [-483]	3068 (+6)
	DFTBA(1.1)	”	3546 (+1)	”	3037 (-25)
	B3LYP/AN07D	3595	3589 (-6)	3215 [-380]	3265 (+50)
3. C-H	DFTBA(0.8/0.9)	3050	3051 (+1)	3049 [-1]	3050 (+1)
	DFTBA(1.1)	”	3052 (+2)	”	3050 (+1)
	B3LYP/AN07D	3253	3254 (+1)	3253 [0]	3254 (+1)
4. C-H	DFTBA(0.8/0.9)	2970	3008 (+38)	2972 [+2]	3006 (+34)
	DFTBA(1.1)	”	2999 (+29)	”	2996 (+24)
	B3LYP/AN07D	3212	3233 (+21)	3214 [+2]	3232 (+18)
5. C=O	DFTBA(0.8/0.9)	1715	1676 (-39)	1704 [-11]	1681 (-23)
	DFTBA(1.1)	”	1681 (-34)	”	1684 (-20)
	B3LYP/AN07D	1806	1749 (-57)	1773 [-33]	1729 (-44)
6. C=O	DFTBA(0.8/0.9)	1649	1540 (-109)	1631 [-18]	1539 (-92)
	DFTBA(1.1)	”	1611 (-38)	”	1611 (-20)
	B3LYP/AN07D	1771	1696 (-75)	1740 [-31]	1699 (-41)

Table 5.10: Solvatochromic shifts in cm^{-1} for selected harmonic frequencies of Uracil in aqueous solution. IEF-PCM model has been used. In round brackets the solvatochromic shifts due to the PCM model (i.e. the difference of Uracil+PCM and Uracil+4H_2O+PCM frequencies, with respect to Uracil and Uracil+4H_2O frequencies, respectively). In square bracket the solvatochromic shifts due to the inclusion of 4 explicit water molecules (i.e. the difference of Uracil+4H_2O frequencies, with respect to Uracil frequencies).

	modes	symm.	exp.[a] (gas-phase)	Uracil	Uracil +PCM	Uracil +4H_2O	Uracil +4H_2O +PCM	exp.[b,c] (water)
1	ν(N-H)	A'	3485	3502 (17)	-	-	-	-
2	ν(N-H)	A'	3435	3459 (24)	-	-	-	-
3	ν(C-H)	A'	3117[d]	3168 (51)	-	-	-	-
4	ν(C-H)	A'	3072[d]	3103 (31)	-	-	-	-
5	ν(C=O)	A'	1764	1746 (17)	1689 (18)	1708 (0)	1660 (47)	1708[b]
6	ν(C=O)	A'	1706	1731 (25)	1642 (28)	1685 (14)	1641 (29)	1671[b]
7	ν(C=C)	A'	1643	1608 (34)	1608 (22)	1601 (29)	1599 (31)	1631[b]
8	δ	A'	1472	1444 (27)	1463 (42)	1476 (29)	1470 (35)	1506[c]
9	δ	A'	1400	1346 (53)	1357 (98)	1438 (17)	1431 (24)	1456[c]
10	δ	A'	1389	1362 (26)	1371 (44)	1385 (30)	1394 (21)	1416[c]
11	δ	A'	1359	1358 (0)	1380 (-)	1376 (-)	1380 (-)	-
12	δ	A'	1217	1187 (29)	1201 (23)	1206 (18)	1220 (4)	1225[c]
13	δ	A'	1185	1182 (2)	1201 (35)	1203 (33)	1209 (27)	1237[b]
14	δ	A'	1075	1068 (6)	1107 (12)	1086 (8)	1084 (10)	1095[b]
15	γ	A"	987	954 (32)	-	-	-	-
16	δ	A'	980	957 (22)	-	-	-	-
17	δ	A'	958	947 (10)	-	-	-	-
18	γ	A"	804	809 (5)	-	-	-	-
19	δ	A'	759	759 (0)	-	-	-	-
20	γ	A"	757	762 (5)	-	-	-	-
21	γ	A"	718	718 (0)	-	-	-	-
22	γ	A"	662	678 (16)	-	-	-	-
23	δ	A'	562	550 (11)	-	-	-	-
24	γ	A"	551	575 (24)	-	-	-	-
25	δ	A'	537	539 (2)	-	-	-	-
26	δ	A'	516	515 (0)	-	-	-	-
27	γ	A"	411	402 (8)	-	-	-	-
28	δ	A'	391	384 (6)	-	-	-	-
29	γ	A"	185	170 (14)	-	-	-	-
30	γ	A"	132[d]	158 (26)	-	-	-	-
MAE(modes 5-10,12-14)				24 (19)[e]	36	20	25	
MAX				53	98	33	47	

Table 5.11: Hybrid DFT/DFTB frequencies in cm^{-1} of Uracil in aqueous solution. Harmonic frequencies at B3LYP/AN07D level, anharmonic shifts at DFTBA level. In parenthesis the absolute errors with respect to experimental frequencies. Errors of Uracil have been computed with respect to gas-phase measurements. Errors of Uracil+4H_2O,Uracil+PCM and Uracil+4H_2O+PCM, have been computed with respect to measurements in aqueous solution. Mean Absolute Errors (MAE) have been computed with reference to experimental frequencies. δ and γ are in plane and out of plane modes respectively. [a] Biczysko et al.[6] and references therein. [b] Experimental frequencies from Raman spectra of Berthier et al.[7] [c] Experimental frequencies from IR spectra of Berthier et al.[7] [d] Not experimentally available (the value of hybrid CC/DFT calculations from Biczysko et al.[6] have been used). [e] MAE in parenthesis refers to all modes.

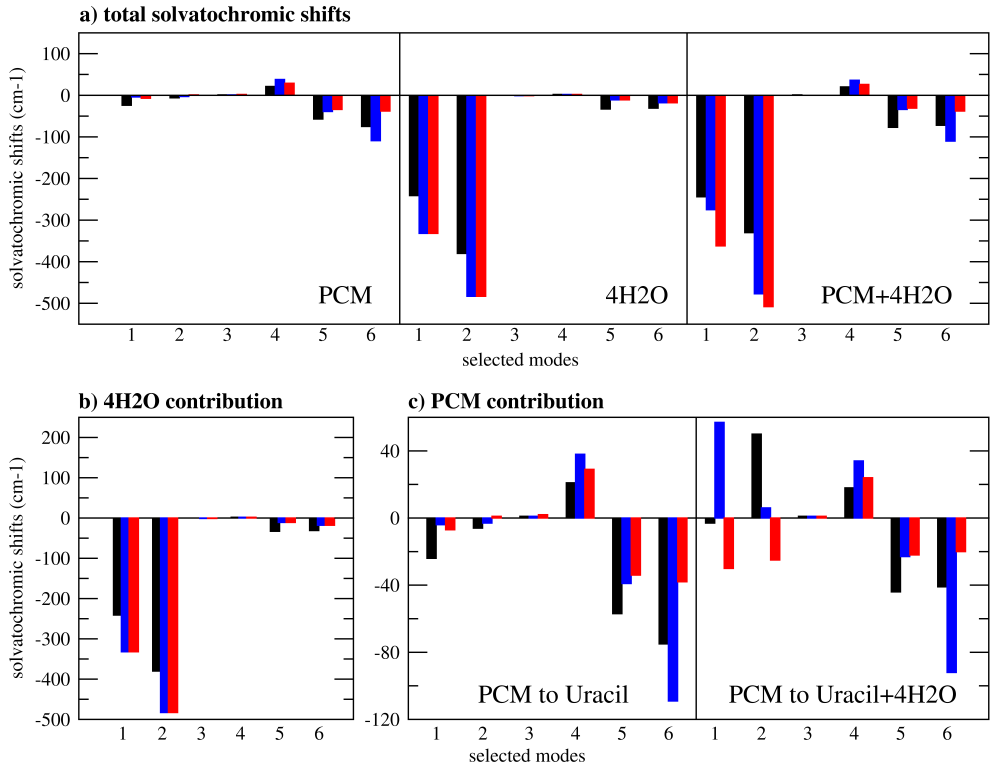


Figure 5.4: Solvatochromic shifts in cm^{-1} of the harmonic frequencies of Uracil in aqueous solution, modelled using different computational approaches. Mode labels on each x-axis, have been defined in Table 5.10. Black bars refer to DFT computations, Blue bars to DFTBA(0.8/0.9), red bars to DFTBA(1.1), see Table 5.10 for the numerical values of the harmonic frequencies. a) Solvatochromic shifts computed w.r.t. the harmonic frequencies of Uracil in gas-phase. b) Contribution to the total solvatochromic shift due to the inclusion of four explicit water molecules (i.e. the difference between the harmonic frequencies of Uracil+$4H_2O$ and Uracil in gas-phase). c) Contribution to the total solvatochromic shift due to the PCM (i.e. the difference between the harmonic frequencies of Uracil+PCM and Uracil+$4H_2O$+PCM w.r.t. the harmonic frequencies of Uracil and Uracil+$4H_2O$, respectively).

5.6 Test cases: electronic excitation energies of Nile Red and Uracil

As a concluding remark, we propose two examples of the application of the time dependent approach to the computation of the solvent effects on the vertical excitation energies.

The Nile Red molecule (shown in Figure 5.5) is a very interesting dye due to its peculiar electronic properties,[238,239] such as significant solvatochromic shifts on the absorption maxima and a dual fluorescence in the emission spectrum. In Table 5.13 the vertical excitation energies of the HOMO-LUMO transition of the Nile Red in different solvents computed with the TD-DFTB/PCM scheme are reported, and compared with other experimental and computational results, available in literature. The vertical excitation energies at the TD-DFTB/PCM level are in quantitative agreement with the experimental energies, being the maximum error of about 0.1 eV in Heptane and Acetonitrile solvents, and just 0.01 eV in Benzene. The accuracy of the TD-DFTB/PCM approach is comparable to the DFT calculations, also when large basis sets are employed, as in the case of the B3LYP/6-311G(d,p). In particular, it is worth of note that using an hybrid scheme in which the B3LYP/6-311G(d,p) method is applied for the calculation of the vertical energies on the DFTB geometries, the results are intermediate between the full DFTB and the full DFT calculations, suggesting that when an high accuracy is required the DFTB/PCM scheme can be combined with the DFT methods. The orbitals involved in the transition of the Nile Red in the gas phase are shown in Figure 5.6, and they suggest that the electronic transition leads to an intramolecular charge transfer, in agreement with other results obtained at DFT level (see Rappon et al.[238] and references therein).

The calculation of the vertical excitation energies of Uracil in aqueous solution is more challenging, since both explicit water molecules and bulk effects need to be considered in order to correctly compute the energies of the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions. In Figure 5.3 the first shell of solvation of Uracil, including four water molecules, has been shown. As reported by Improta et al.,[91] the solvatochromic shifts of such transitions are significantly affected by either specific interactions between the water molecules and the lone pairs of the Oxygen atoms of Uracil, and by the electrostatic interaction between the bulk and the dipole moments of the ground and excited states. For these reason gas phase calculations usually lead to an incorrect ordering of the excited states with respect to the experiments. Thus, this system is an interesting test case in order to discuss the performances of our TD-DFTB/PCM approach. We first optimized the geometries of Uracil in gas-phase and with different solvation models (Uracil+PCM, Uracil+4H_2O and Uracil+4H_2O+PCM), using the DFTBA method, and default cavities for PCM calculations, then we computed the transition energies at DFTBA(1.1) level. Also single point calculations using the reduced cavity scaling factor ($\alpha = 0.8/0.9$),

pure DFT methods (PBE/STO-3G and PBE-AVTZ) and hybrid DFT methods (B3LYP/AVTZ, PBE0/AVTZ) have been performed on the same geometries, in order to compare the DFTBA results with more accurate methods. In Table 5.12 the transition energies of the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions are reported, along with accurate results obtained at PBE0 level with extended basis sets,[91] for which discrepancies lower than 0.1 eV have been found with respect to experimental results (see the original work for details). When hybrid functionals are used in conjunction with the most accurate solvent model (Uracil+4H_2O+PCM), the inversion of the first and second excited state is observed, in agreement with the results reported in literature. Such an inversion is due the strong blue-shift of the $n \rightarrow \pi^*$ transition (about 0.5 eV), while the energy of the $\pi \rightarrow \pi^*$ is slightly affected by the solvent (red shift of less than 0.2 eV). since the variation of the dipole moment is very small. Using pure DFT functionals (see the PBE/AVTZ results), we note similar effects on the transition energies, with a sensible blue shift of the $n \rightarrow \pi^*$ transition (0.45 eV), and a very small variation of the energy associated to the $\pi \rightarrow \pi^*$. However, although the solvatochromic shifts are in quantitative agreement with the hybrid functionals, the total transition energies are quite different, due to a bias error affecting the gas phase energies. For this reason we do not observe any inversion of the stability of the excited states in water. In DFTBA calculations, from a qualitative point of view, both the cavity scaling factors lead to similar results, showing the blue-shift of the $n \rightarrow \pi^*$ transition and very small variations of the energy of the $\pi \rightarrow \pi^*$ transition. From a quantitative point of view, the default cavity leads to an underestimation of the solvatochromic shift of the $n \rightarrow \pi^*$ (0.33 eV), while the smaller cavity reproduces it in quantitative agreement with reference results (0.54 eV). This is due to the enhanced solute-solvent electrostatic interaction, and is in line with the discussion about the energetics of solvation at the beginning of this section. In analogy with the PBE/AVTZ calculations, the correct ordering of the excited states is not observed because of the underestimation of the energy of the $n \rightarrow \pi^*$ transition in the gas phase, while the solvation effects are very well computed employing the Uracil+4H_2O+PCM solvation model. Finally, it is worth to note that also the oscillator strenghts computed by the DFTBA(0.8/0.9) method are very similar to the ones computed at PBE/AVTZ level. So, as a final remark we can summarize that in the proposed test cases the TD-DFTB/PCM method shows an accuracy higher than the CIS-HF/PCM and the TD-PBE/STO-3G approaches, resembling the results of a pure DFT method employing a large basis set.

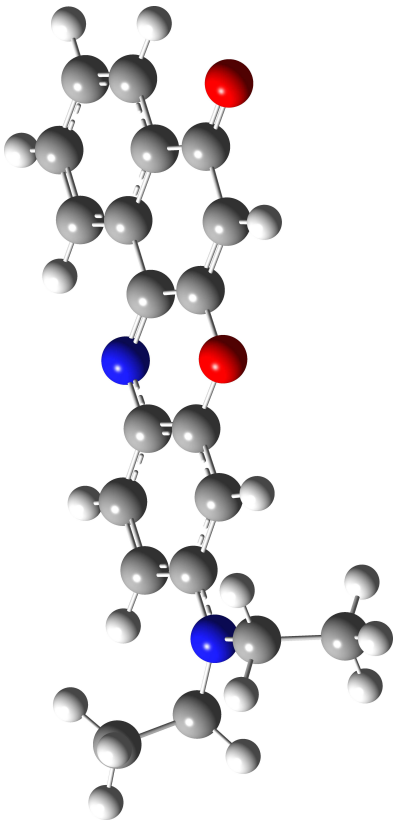


Figure 5.5: The most stable conformer of Nile Red.[238,239]

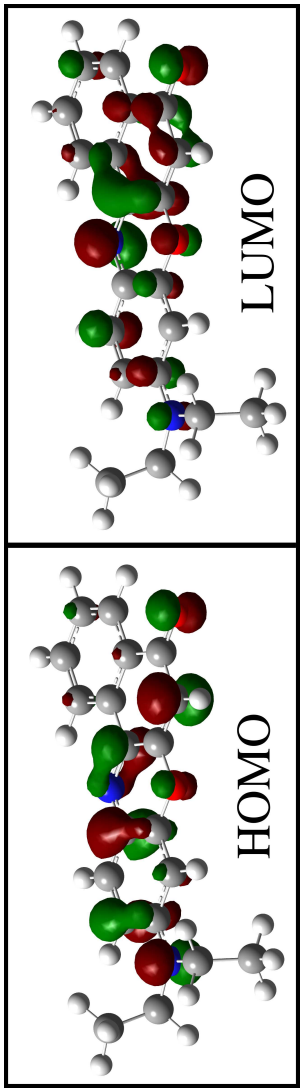


Figure 5.6: HOMO and LUMO molecular orbitals of Nile Red, as computed at DFTB level in this work.

		gas-phase	Uracil +PCM	Uracil +4H_2O	Uracil +4H_2O +PCM
	DFTBA(1.1)	3.64 (0.00)	3.90 (0.00)	3.79 (0.00)	3.97 (0.00)
	DFTBA(0.8/0.9)	3.64 (0.00)	4.18 (0.00)	3.79 (0.00)	4.18 (0.00)
	PBE/STO-3G	3.30 (0.00)	3.23 (0.00)	3.32 (0.00)	3.40 (0.00)
$n \rightarrow \pi^*$	PBE/AVTZ	3.85 (0.00)	4.10 (0.00)	4.11 (0.00)	4.30 (0.00)
	B3LYP/AVTZ	4.53 (0.00)	4.79 (0.00)	4.95 (0.00)	5.00 (0.00)
	PBE0/AVTZ	4.65 (0.00)	4.90 (0.00)	4.94 (0.00)	5.11 (0.00)
	PBE0[a]	4.80 (0.00)	5.09 (0.00)	4.97 (0.00)	5.28 (0.00)
	DFTBA(1.1)	4.73 (0.03)	4.76 (0.07)	4.66 (0.03)	4.74 (0.08)
	DFTBA(0.8/0.9)	4.73 (0.03)	4.82 (0.10)	4.66 (0.03)	4.78 (0.10)
	PBE/STO-3G	4.90 (0.01)	4.81 (0.02)	4.43 (0.00)	4.58 (0.01)
$\pi \rightarrow \pi^*$	PBE/AVTZ	4.60 (0.05)	4.56 (0.09)	4.52 (0.02)	4.60 (0.10)
	B3LYP/AVTZ	5.01 (0.11)	4.95 (0.16)	5.01 (0.07)	4.93 (0.17)
	PBE0/AVTZ	5.13 (0.13)	5.05 (0.18)	5.06 (0.14)	5.03 (0.19)
	PBE0[a]	5.26 (0.14)	5.17 (0.19)	5.26 (0.14)	5.16 (0.20)

Table 5.12: Vertical excitation energies (eV) for $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ electronic transitions of Uracil molecule in aqueous solution. All geometries optimized at DFTBA level, with $\alpha = 1.1$ when PCM is applied. C-PCM model has been used for consistency with reference DFT caculations. The difference between C-PCM and IEF-PCM vertical energies are on the third decimal digits. Oscillator strenghts have been reported in parenthesis. Assignment of transitions have been checked by comparison with the oscillator strenghts reported in literature and by visual inspection of molecular orbitals. [a] PBE0/6-311+G(2d,2p) energies on PBE0/6-31+G(d,p) geometries, taken from Improta et al.[91]

solvent	dielectric constant	TD-DFTBA	TD-B3LYP/[a] 6-311G(d,p)	TD-B3LYP/ 6-311G(d,p)	TD-B3LYP[b] 6-31+G(d)	TD-PBE0[c] 6-311G(d,p)	exp.
gas-phase	-	2.47	2.66	2.74[d]	2.68	2.83	-
heptane	1.9113	2.32	2.49	2.57[d]	-	2.65	2.43[c]
benzene	2.2706	2.30	2.46	2.54	2.47	-	2.31[b]
acetonitrile	35.688	2.25	2.41	2.48	2.40	-	2.14[b]

Table 5.13: Vertical excitation energies (eV) corresponding to the HOMO-LUMO transition of Nile Red (see Figure 5.5 and 5.6), in different solvents. [a] TD-B3LYP/6-311G(d,p) energies on TD-DFTBA geometries. [b] Rappon et al.[238] and references therein. [c] Mennucci al.[239] and references therein. [d] Agreement within 0.01 eV with respect to other calculations at the same level of theory.[239]

Chapter 6

Conclusions

In this thesis the basic framework for the computations of spectroscopical quantities for large systems has been presented. As shown in Chapter 1 the VPT2 approach is a flexible and reliable tool for the computation of the Infrared spectra, and in general of vibrational properties. Using coupled-cluster methods, or hybrid coupled-cluster and DFT approaches the computed vibrational properties are always in nearly quantitative agreement with the experimental data. When the dimension of the systems increases, the RD-VPT2 model can be used as a first chance, allowing the simulation of the spectra using fully coherent DFT force fields. The RD-VPT2 approach is effective especially for those systems which can be partitioned in fragments, and the normal modes of the different fragments are not strongly coupled. This was the case of the adsorption process of the glycine molecule adsorbed on the (100) Silicon surface, for which the particular morphology of the system induced a weak coupling between the normal modes of the molecule and the surface phonons. Such an approach has been demonstrated to be more reliable than a VSCF approach, applied to the same system, in which a semi-empirical PM3 PES was used for the representation of the potential. The main drawback of the RD-VPT2 approach is that it is system-dependent, since it fails when strong couplings between the normal modes are found.

In order to extend the computations of the anharmonic spectra to larger systems the DFTB method has been proposed, within the framework of multi-layer approaches, and the TD-DFTB, DFTB/PCM and TD-DFTB/PCM schemes have been implemented. In the last chapters, we have reported the essential details of a new fully coherent and numerically efficient implementation of the DFTB/PCM and TD-DFTB/PCM schemes including energies for ground and excited electronic states together with analytical gradients and Hessians for ground states. From the one side, the numerical differentiation of the analytical Hessian gives access to anharmonic computations in solution, within the latest effective and stable generalizations of the second order vibrational perturbation theory. From the other side, the availability of the TD-DFTB/PCM allows the study of photophysical and photochemical processes involving large species in solution. Since the accuracy of the DFTB/PCM and TD-

DFTB/PCM results rivals that of the DFT model from which they were derived, we suggest that further improvements can be achieved by a reparametrization of the method with respect to more accurate references. Already at the present stage, coupling our approach with a more refined QM model for a reduced part of the system and/or lower-order Taylor series expansion of the potential energy surface (e.g. harmonic vs. anharmonic approximation) provides results of high accuracy. Although further developments are surely in order along those and related routes (e.g. TD-DFTB/PCM analytical gradients) we think that our implementation already represents a new useful tool for the computational spectroscopy studies of large systems in condensed phases.

All these results are encouraging towards effective and reliable computations of vibrational and electronic properties with multi-layered schemes, of the type QM/QM'/PCM, QM/QM'/MM/PCM.

Appendix A

Spectral lineshapes

The *molar absorption coefficient* can be written as

$$\varepsilon(\nu) = \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc}\nu \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \delta(\nu_{IF} - \nu), \tag{A.1}$$

where the relevant quantities have been defined in Chapter 1. Since in Eq. A.1 the ν frequency does not depend on the initial and final states, it can be moved inside the summation, and the integration over the frequency range $[\nu_1, \nu_2]$ can be easily accomplished, obtaining the *integrated cross section*

$$G_{\nu_1,\nu_2} = \int_{\nu_1}^{\nu_2} \varepsilon(\nu) d\nu = \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc} \sum_{\substack{IF \\ \nu_1<\nu_{IF}<\nu_2}} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \nu_{IF} = \sum_{\substack{IF \\ \nu_1<\nu_{IF}<\nu_2}} G_{IF}, \tag{A.2}$$

where the G_{IF} is defined as

$$G_{IF} = \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \nu_{IF}. \tag{A.3}$$

The spectral lineshapes of an IR spectrum can be obtained[240] from the knowledge of the G_{IF}s. Choosing a Gaussian function $\mathcal{G}_\gamma(\nu)$ as a convoluting function, of full width at half maximum given by γ

$$\mathcal{G}_\gamma(\nu) = \frac{1}{\gamma}\sqrt{\frac{4\ln 2}{\pi}} \exp\left[-4\ln 2 \left(\frac{\nu}{\gamma}\right)^2\right]. \tag{A.4}$$

Then, the convolution of $\varepsilon(\nu)$ with $g_\gamma(\nu)$ is

$$\begin{aligned}
\varepsilon_\gamma(\nu) =& \varepsilon(\nu') * \mathcal{G}_\gamma(\nu' - \nu) \\
=& \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc} \frac{1}{\gamma}\sqrt{\frac{4\ln 2}{\pi}} \int d\nu' \nu' \exp\left[-4\ln 2 \left(\frac{\nu' - \nu}{\gamma}\right)^2\right] \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \delta(\nu_{IF} - \nu') \\
=& \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc} \frac{1}{\gamma}\sqrt{\frac{4\ln 2}{\pi}} \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \int d\nu' \nu' \exp\left[-4\ln 2 \left(\frac{\nu' - \nu}{\gamma}\right)^2\right] \delta(\nu_{IF} - \nu') \\
=& \frac{4N_A\pi^2}{3\cdot 10^3 \ln(10)hc} \frac{1}{\gamma}\sqrt{\frac{4\ln 2}{\pi}} \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \nu_{IF} \exp\left[-4\ln 2 \left(\frac{\nu_{IF} - \nu}{\gamma}\right)^2\right],
\end{aligned} \tag{A.5}$$

and using the definitions of Eq. A.3 and A.4

$$\varepsilon_\gamma(\nu) = \sum_{IF} G_{IF}\mathcal{G}_\gamma(\nu_{IF} - \nu). \tag{A.6}$$

Analogously, the same approach can be applied choosing a Lorentzian function of the type

$$\mathcal{L}_\gamma(\nu) = \left(\frac{2}{\pi\gamma}\right) \frac{1}{1 + 4\left(\frac{\nu}{\gamma}\right)^2} \tag{A.7}$$

The convolution of $\varepsilon(\nu)$ with $\mathcal{L}_\gamma(\nu)$ is

$$\begin{aligned}
\varepsilon_\gamma(\nu) =& \varepsilon(\nu') * \mathcal{L}_\gamma(\nu' - \nu) \\
=& \frac{4N_A\pi^2}{3 \cdot 10^3 \ln(10) hc} \left(\frac{2}{\pi\gamma}\right) \int d\nu'\nu' \frac{1}{1 + 4\left(\frac{\nu'-\nu}{\gamma}\right)^2} \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \delta(\nu_{IF} - \nu') \\
=& \frac{4N_A\pi^2}{3 \cdot 10^3 \ln(10) hc} \left(\frac{2}{\pi\gamma}\right) \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \int d\nu'\nu' \frac{1}{1 + 4\left(\frac{\nu'-\nu}{\gamma}\right)^2} \delta(\nu_{IF} - \nu') \\
=& \frac{4N_A\pi^2}{3 \cdot 10^3 \ln(10) hc} \left(\frac{2}{\pi\gamma}\right) \sum_{IF} N_I |\langle \boldsymbol{\mu} \rangle_{IF}|^2 \nu_{IF} \frac{1}{1 + 4\left(\frac{\nu_{IF}-\nu}{\gamma}\right)^2},
\end{aligned} \tag{A.8}$$

and using the definitions of Eq. A.3 and A.7

$$\varepsilon_\gamma(\nu) = \sum_{IF} G_{IF}\mathcal{L}_\gamma(\nu_{IF} - \nu). \tag{A.9}$$

It is worth to note that γ is an empirical parameter, and we have assumed that all the transitions have the same full width at half maximum.

Appendix B

Atomic quantities

Eq. 4.33, 4.37, 4.41 and 4.42 have been used in order to compute the relevant DFTB quantities for simple atomic systems. In Table B.1 the total and spin charges, the Fock matrix elements and the total energies have been reported for ^{2}H, ^{1}H$^{(+1)}$, ^{1}H$^{(-1)}$, ^{5}C and ^{4}C$^{(-1)}$. All the quantities have been expressed using only the parameters employed by the DFTB method, and calculations with the GAUSSIAN[133] code have been performed in order to verify the formulas with the numerical values. As a general remark, for all the systems the off-diagonal Fock matrix elements of the type F_{sp} are zero due to the zero onsite values of the corresponding interaction curves and to the zero off-diagonal elements of the atomic Overlap matrix. In fact, $S_{\mu\nu} = 0$ if $\mu \neq \nu$ and both are centered on the same atom. Then, in all cases the eigenvalues ε required by the evaluation of the energy coincides with the corresponding Fock matrix elements. The notation ε has been used anyway since in general for molecular systems the Fock matrix is not diagonal.

Moreover, when $\mathbf{P}^\beta = 0$ (see ^{2}H, ^{5}C) the Fock matrix $\mathbf{F}^\beta$ is defined following Eq. 4.41, but the related energy contribution (second term of Eq. 4.42) is strictly zero since it is calculated as the trace with the β density matrix.

	$^{2}\mathrm{H}$	$^{1}\mathrm{H}^{(+1)}$	$^{1}\mathrm{H}^{(-1)}$	$^{5}\mathrm{C}$	$^{4}\mathrm{C}^{(-1)}$
Z_A	$+1$	$+1$	$+1$	$+4$	$+4$
q_A	$P^{\alpha}_{ss}S_{ss} = +1$	0	$(P^{\alpha}_{ss} + P^{\beta}_{ss})S_{ss} = +2$	$P^{\alpha}_{ss}S_{ss} + P^{\alpha}_{xx}S_{xx} + P^{\alpha}_{yy}S_{yy} + P^{\alpha}_{zz}S_{zz} = +4$	$(P^{\alpha}_{ss} + P^{\beta}_{ss})S_{ss} + P^{\alpha}_{xx}S_{xx} + P^{\alpha}_{yy}S_{yy} + P^{\alpha}_{zz}S_{zz} = +5$
Δq_A	0	$+1$	-1	0	-1
p_s	$P^{\alpha}_{ss}S_{ss} = +1$	0	$(P^{\alpha}_{ss} - P^{\beta}_{ss})S_{ss} = 0$	$P^{\alpha}_{ss}S_{ss} = +1$	$(P^{\alpha}_{ss} - P^{\beta}_{ss})S_{ss} = 0$
p_p	0	0	0	$P^{\alpha}_{xx}S_{xx} + P^{\alpha}_{yy}S_{yy} + P^{\alpha}_{zz}S_{zz} = +3$	$P^{\alpha}_{xx}S_{xx} + P^{\alpha}_{yy}S_{yy} + P^{\alpha}_{zz}S_{zz} = +3$
F^{α}_{ss}	$h_s + W_{ss}$	$h_s - \gamma_{HH}$	$h_s + \gamma_{HH}$	$h_s + W_{ss} + 3W_{sp}$	$h_s + 3W_{sp} + \gamma_{CC}$
F^{β}_{ss}	$h_s - W_{ss}$	$h_s - \gamma_{HH}$	$h_s + \gamma_{HH}$	$h_s - W_{ss} - 3W_{sp}$	$h_s - 3W_{sp} + \gamma_{CC}$
F^{α}_{sp}	0	0	0	0	0
F^{β}_{sp}	0	0	0	0	0
F^{α}_{pp}	0	0	0	$h_p + W_{sp} + 3W_{pp}$	$h_p + 3W_{pp} + \gamma_{CC}$
F^{β}_{pp}	0	0	0	$h_p + W_{sp} + 3W_{pp}$	$h_p - 3W_{pp} + \gamma_{CC}$
E	$\varepsilon^{\alpha}_{s} - \frac{1}{2}W_{ss}$	$\frac{1}{2}\gamma_{HH}$	$^{**}2\varepsilon^{\alpha}_{s} + \frac{3}{2}\gamma_{HH}$	$\varepsilon^{\alpha}_{s} + 3\varepsilon^{\alpha}_{p} + \frac{1}{2}\left(W_{ss} + 9W_{pp} + 6W_{sp}\right)$	$\varepsilon^{\alpha}_{s} + \varepsilon^{\beta}_{s} + 3\varepsilon^{\alpha}_{p} + \frac{1}{2}\left(9\gamma_{CC} + 9W_{pp}\right)$

Table B.1: In the following table h indicates the onsite values of the Fock interaction curves introduced in Section 4.1.4, while ε refers to the occupied eigenvalues of the Fock matrix. For the open-shell systems α spin has been given to the unpaired electrons. $[^{**}]$ $\varepsilon^{\alpha}_{s} = \varepsilon^{\beta}_{s}$.

Appendix C

Integral evaluation

It is well known that linear combinations of Gaussian-type functions can reproduce very well the trend of the atomic orbitals χ_μ. As a consequence the integrals in both sides of Eq. 4.77 can be evaluated in terms of Gaussian distributions for the electronic densities χ_μ and χ_ν. The integral in the right hand side of Eq. 4.77 is then a two-center integral involving a Gaussian function with exponent α located on atom A, and a Gaussian function with exponent β located on atom B (neglecting the normalization coefficients). This integral has an analytical known solution

$$\int d\vec{r}' e^{-\alpha|\vec{r}'-\vec{R_A}|^2} \vec{r}' e^{-\beta|\vec{r}'-\vec{R_B}|^2} = \left(\frac{\pi}{\alpha+\beta}\right)^{\frac{3}{2}} \exp\left(-\frac{\alpha\beta}{\alpha+\beta}|\vec{R_A}-\vec{R_B}|^2\right)\left(\frac{\alpha\vec{R_A}+\beta\vec{R_B}}{\alpha+\beta}\right) \quad \text{(C.1)}$$

where α and β in principle are different. In the right hand side of Eq. 4.77 we have the Overlap integral, which can be analogously estimated

$$\int d\vec{r}' e^{-\alpha|\vec{r}'-\vec{R_A}|^2} e^{-\beta|\vec{r}'-\vec{R_B}|^2} = \left(\frac{\pi}{\alpha+\beta}\right)^{\frac{3}{2}} \exp\left(-\frac{\alpha\beta}{\alpha+\beta}|\vec{R_A}-\vec{R_B}|^2\right) \quad \text{(C.2)}$$

Substituting these two expressions in Eq. 4.77 we find that the computation of the dipole moment using the density of Eq. 4.62 rather than Eq. 4.47 is equivalent to use fixed, unitary values for the exponents of the Gaussian functions.

$$\left(\frac{\alpha\vec{R_A}+\beta\vec{R_B}}{\alpha+\beta}\right) \simeq \left(\frac{\vec{R_A}+\vec{R_B}}{2}\right) \quad \text{(C.3)}$$

We can also estimate the integrals of the electrostatic potential using Gaussian distributions

$$\vec{R_P} = \frac{\alpha\vec{R_A}+\beta\vec{R_B}}{\alpha+\beta}$$

$$\int d\vec{r}' \frac{e^{-\alpha|\vec{r}'-\vec{R_A}|^2} e^{-\beta|\vec{r}'-\vec{R_B}|^2}}{|\vec{r}-\vec{r}'|} = \left(\frac{\pi}{\alpha+\beta}\right)^{\frac{3}{2}} \exp\left(-\frac{\alpha\beta}{\alpha+\beta}|\vec{R_A}-\vec{R_B}|^2\right) \frac{\mathrm{erf}\left(\sqrt{(\alpha+\beta)|\vec{r}-\vec{R_P}|^2}\right)}{|\vec{r}-\vec{R_P}|^2} \quad \text{(C.4)}$$

Eq. 4.66 then reduces to

$$\frac{\mathrm{erf}\left(\sqrt{(\alpha+\beta)|\vec{r}-\vec{R_P}|^2}\right)}{|\vec{r}-\vec{R_P}|^2} \simeq \frac{1}{2}\left(\frac{1}{|\vec{r}-\vec{R_A}|}+\frac{1}{|\vec{r}-\vec{R_B}|}\right) \tag{C.5}$$

Which is formally exact when the point $\vec{r}$ at which the potential is evaluated is far from the location of the sources.

Bibliography

[1] Tasinato, N.; Pietropolli Charmet, A.; Stoppa, P.; Giorgianni, S.; Buffa, G. *J. Chem. Phys.* **2010**, *132*, 044315.

[2] Stoppa, P.; Pietropolli Charmet, A.; Tasinato, N.; Giorgianni, S.; Gambi, A. *J. Phys. Chem. A* **2009**, *113*, 1497.

[3] Pietropolli Charmet, A.; Stoppa, P.; Tasinato, N.; Baldan, A.; Giorgianni, S.; Gambi, A. *J. Chem. Phys.* **2010**, *133*, 044310.

[4] Espinoza, C.; Szczepanski, J.; Vala, M.; Polfer, N. *J. Phys. Chem. A* **2010**, *114*, 5919.

[5] Stepanian, S.; Reva, I.; Radchenko, E.; Rosado, M.; Duarte, M.; Fausto, R.; Adamowicz, L. *J. Phys. Chem. A* **1998**, *102*, 1041.

[6] Puzzarini, C.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 3702.

[7] Aamouche, A.; Ghomi, M.; Coulombeau, C.; Jobic, H.; Grajcar, L.; Baron, M.; Baumruk, V.; Turpin, P.; Henriet, C.; Berthier, G. *J. Phys. Chem.* **1996**, *100*, 5224.

[8] Lafosse, A.; Bertin, M.; Domaracka, A.; Plizska, D.; Illenberger, E.; Azria, R. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5564.

[9] E. Bright Wilson, J., Decius, J. C., Cross, P. C., Eds. *MOLECULAR VIBRATION. The Theory of Infrared and Raman Vibrational Spectra;* McGRAW-HILL BOOK COMPANY, INC.: New York, USA, 1955.

[10] Califano, S., Ed. *Vibrational States;* John Wiley & Sons: J. W. Arrowsmith Ltd., Winterstoke Road, Bristol BS2 2NT, 1976.

[11] Barone, V., Ed. *Computational Strategies for Spectroscopy, from Small Molecules to Nano Systems.;* John Wiley & Sons, Inc. Hoboken: New Jersey, 2011.

[12] Mills, I. M. In Rao, K. N., Mathews, C. W., Eds., *Molecular Spectroscopy: Modern Research, Volume 1*, page 115, 1972.

[13] Nielsen, H. H. *Rev. Mod. Phys.* **1951**, *23*, 90.

[14] Aliev, M.; Watson, J. In *Molecular Spectroscopy: Modern Research*; Rao, K., Ed.; Academic Press, 1985; pages 1 – 67.

[15] Papoušek, D. D.; Aliev, M. R. (Mamed Ragimovich), . *Molecular vibrational-rotational spectra : theory and applications of high resolution infrared, microwave, and Raman spectroscopy of polyatomic molecules;* Amsterdam ; New York : Elsevier Scientific Pub. Co. ; New York : Distribution for the U.S.A. and Canada, Elsevier /North-Holland, 1982.

[16] Kongsted, J.; Christiansen, O. *J. Chem. Phys.* **2006**, *125*, 124108.

[17] Chaban, G. M.; Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1999**, *111*, 1823.

[18] Chaban, G.; Jung, J.; Gerber, R. *J. Phys. Chem. A* **2000**, *104*, 10035.

[19] Suwan, I.; Gerber, R. B. *Chem. Phys.* **2010**, *373*, 267.

[20] Shemesh, D.; Mullin, J.; Gordon, M.; Gerber, R. *Chem. Phys.* **2008**, *347*, 218.

[21] Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1996**, *105*, 10332.

[22] Amos, R. D.; Handy, N. C.; Green, W. H.; Jayatilaka, D.; Willets, A.; Palmieri, P.

[23] Barone, V. *J. Chem. Phys.* **2004**, *120*, 3059.

[24] Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108.

[25] Bloino, J.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 1015.

[26] Martin, J. M. L.; Lee, T. J.; Taylor, P. R.; Francois, J. P. *J. Chem. Phys.* **1995**, *103*, 2589.

[27] Stanton, J. F.; Gauss, J.

[28] Vázquez, J.; Stanton, J. F. *Mol. Phys.* **2006**, *104*, 377.

[29] Barone, V.; Biczysko, M.; Bloino, J. *Phys. Chem. Chem. Phys. just accepted.*

[30] Bowman, J. M. *Acc. Chem. Res.* **1986**, *19*, 202.

[31] Bowman, J. M. *Science* **2000**, *290*, 724.

[32] Gerber, R. B.; Rattner, M. *Adv. Chem. Phys.* **1988**, *70*, 97.

[33] Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1996**, *105*, 10682.

[34] Luis, J. M.; Torrent-Sucarrat, M.; Christiansen, O.; Kirtman, B. *J. Chem. Phys.* **2007**, *127*, 084118.

[35] Christiansen, O. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942.

[36] Norris, L. S.; Ratner, M. A.; Roitberg, A. E.; Gerber, R. B. *J. Chem. Phys.* **1996**, *105*, 11261.

[37] Christiansen, O. *J. Chem. Phys.* **2003**, *119*, 5773.

[38] Bowman, J. M.; Carter, S.; Huang, X. *Int. Rev. Phys. Chem.* **2003**, *22*, 533.

[39] Carter, S.; Sharma, A. R.; Bowman, J. M.; Rosmus, P.; Tarroni, R. *J. Chem. Phys.* **2009**, *131*, 224106.

[40] Christiansen, O. *J. Chem. Phys.* **2004**, *120*, 2149.

[41] Barone, V.; Festa, G.; Grandi, A.; Rega, N.; Sanna, N. *Chem. Phys. Lett.* **2004**, *388*, 279.

[42] Boese, A. D.; Martin, J. M. L. *J. Phys. Chem. A* **2004**, *108*, 2085.

[43] Kuhler, K. M.; Truhlar, D. G.; Isaacson, A. D. *J. Chem. Phys.* **1996**, *104*, 4664.

[44] Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Catak, S.; Brooks, B. R.; Van Neck, D.; Waroquier, M. *J. Comp. Chem.* **2010**, *31*, 994.

[45] France, S.; Guerin, D. J.; Miller, S. J.; Lectka, T. *Chem. Rev.* **2003**, *103*, 2985.

[46] Fessenden, R. J. *Organic. Chem.* **1990**.

[47] Ghysels, A.; Van Neck, D.; Waroquier, M. *J. Chem. Phys.* **2007**, *127*, 164108.

[48] Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. *J. Chem. Phys.* **2007**, *126*, 224102.

[49] Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Van Neck, D.; Brooks, B. R.; Waroquier, M. *J. Chem. Theory Comput.* **2009**, *5*, 1203.

[50] Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. *Int. J. Quant. Chem.* **2012**, *112*, 2185.

[51] Carnimeo, I.; Biczysko, M.; Bloino, J.; Barone, V. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16713.

[52] Martin, J. M. L.; Taylor, P. R. *Chem. Phys. Lett.* **1996**, *248*, 336.

[53] Martin, J. M. L.; Taylor, P. R.; Lee, T. J. *Chem. Phys. Lett.* **1997**, *275*, 414.

[54] Lee, T. J.; Martin, J. M. L.; Taylor, P. R. *J. Chem. Phys.* **1995**, *102*, 254.

[55] Handy, N. C.; Willetts, A. *Spectrochim. Acta A* **1997**, *53*, 1169.

[56] Watts, J. D.; Gwaltney, S. R.; Bartlett, R. J. *J. Chem. Phys.* **1996**, *105*, 6979.

[57] Del Bene, J. E.; Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1997**, *106*, 6051.

[58] study of vibronic spectra, T.; photodissociation pathways of methane. *J. Chem. Phys.* **1997**, *106*, 2612.

[59] Roos, B. O.; Andersson, K.; Fülscher, M. P. *Chem. Phys. Lett.* **1992**, *192*, 5.

[60] van Harrevelt, R. *J. Chem. Phys.* **2006**, *125*, 124302.

[61] Tajti, A.; Szalay, P.; Császár, A.; Kállay, M.; Gauss, J.; Valeev, E.; Flowers, B.; Vázquez, J.; Stanton, J. *J. Chem. Phys.* **2004**, *121*, 11599.

[62] Bomble, Y.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P.; Császár, A.; Gauss, J.; Stanton, J. *J. Chem. Phys.* **2006**, *125*, 064108.

[63] Harding, M. E.; Vazquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2008**, *128*, 114111.

[64] Harding, M. E.; Vazquez, J.; Gauss, J.; Stanton, J. F.; Kállay, M. *J. Chem. Phys.* **2011**, *135*, 044513.

[65] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

[66] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

[67] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108/1–16.

[68] Neese, F.; Schwabe, T.; Grimme, S. *J. Chem. Phys.* **2007**, *126*, 124115.

[69] Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V. *J. Chem. Theory Comput.* **2010**, *6*, 2115.

[70] Carnimeo, I.; Puzzarini, C.; Tasinato, N.; Stoppa, P.; Charmet, A. P.; Biczysko, M.; Cappelli, C.; Barone, V. *J. Chem. Phys.* **2013**, *139*, 074310.

[71] Biczysko, M.; Bloino, J.; Carnimeo, I.; Panek, P.; Barone, V. *J. Mol. Struct.* **2012**, *1009*, 74.

[72] Carbonniere, P.; Barone, V. *Chem. Phys. Lett.* **2004**, *399*, 226.

[73] Charmet, A. P.; Stoppa, P.; Tasinato, N.; Giorgianni, S.; Barone, V.; Biczysko, M.; Bloino, J.; Cappelli, C.; Carnimeo, I.; Puzzarini, C. *J. Chem. Phys.* **2013**, *139*, 164302.

[74] Biczysko, M.; Bloino, J.; Brancato, G.; Cacelli, I.; Cappelli, C.; Ferretti, A.; Lami, A.; Monti, S.; Pedone, A.; Prampolini, G.; Puzzarini, C.; Santoro, F.; Trani, F.; Villani, G. *Theor. Chem. Acc.* **2012**, *131*, 1201.

[75] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1358.

[76] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *J. Chem. Theory Comput.* **2013**, *9*, 1533.

[77] Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10094.

[78] Grimme, S., Ed. *Calculation of the Electronic Spectra of Large Molecules, Reviews in Computational Chemistry;* John Wiley & Sons, Inc., 2004.

[79] Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12404.

[80] Vreven, T.; Morokuma, K. *J. Comp. Chem.* **2000**, *21*, 1419.

[81] Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760.

[82] Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; M. J. Frisch, M. J. *J. Phys. Chem. B* **2004**, *108*, 4210.

[83] Barone, V.; Biczysko, M.; Brancato, G. *Adv. Quantum Chem.* **2010**, *59*, 17.

[84] Pedone, A.; Biczysko, M.; Barone, V. *ChemPhysChem.* **2010**, *11*, 1812.

[85] Murugan, N.; Kongsted, J.; Rinkevicius, Z.; Aidas, K.; Agren, H. *J. Phys. Chem. B* **2010**, *114*, 13349.

[86] Pedone, A.; Prampolini, G.; Monti, S.; Barone, V. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16689.

[87] Lipparini, F.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 3711.

[88] Lipparini, F.; Cappelli, C.; Barone, V. *J. Chem. Theory Comput.* **2012**, *8*, 4153.

[89] Curutchet, C.; Munoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838.

[90] Steindal, A.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027.

[91] Improta, R.; Barone, V. *J. Am. Chem. Soc.* **2004**, *126*, 14320.

[92] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

[93] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comp. Chem.* **2003**, *24*, 669.

[94] Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *J. Chem. Phys. Lett.* **1996**, *255*, 327.

[95] Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

[96] Cances, E.; Mennucci, B. *Journal of Mathematical Chemistry* **1998**, *23*, 309.

[97] Mennucci, B.; Cances, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506.

[98] Cappelli, C.; Corni, S.; Cammi, R.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2000**, *113*, 11270.

[99] Cappelli, C.; Lipparini, F.; Bloino, J.; Barone, V. *J. Chem. Phys.* **2011**, *135*, 104505.

[100] Mennucci, B.; Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1998**, *109*, 2798.

[101] Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

[102] Pople, J. A.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S136.

[103] Pople, J. A.; Segal, G. A. *J. Chem. Phys.* **1966**, *44*, 3289.

[104] Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. *J. Chem. Phys.* **1967**, *47*, 2026.

[105] Dewar, M.; Zoebisch, E.; Healy, E.; Stewart, J. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

[106] Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173.

[107] Saatzer, P.; Koob, R.; Gordon, M. *J. Chem. Soc., Faraday Trans. 2* **1977**, *2*, 73.

[108] Salahub, D. R.; c. Sandorfy. *Theor. Chem. Acc.* **1971**, *20*, 227.

[109] Neugebauer, J. *ChemPhysChem.* **2009**, *10*, 3148.

[110] Jalkanen, K. J.; Jurgensen, V. W.; Claussen, A.; Rahim, A.; Jensen, G. M.; Wade, R. C.; Nardi, F.; Jung, C.; Degtyarenko, I. M.; Nieminen, R. M.; Herrmann, F.; Knapp-Mohammady, M.; Niehaus, T. A.; Frimand, K.; Suhai, S. *Int. J. Quant. Chem.* **2006**, *106*, 1160.

[111] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

[112] Zheng, G.; Lundberg, M.; Jakowski, J.; Vreven, T.; Frisch, M. J.; Morokuma, K. *Int. J. Quant. Chem.* **2009**, *109*, 1841.

[113] Harris, J. *Phys. Rev. B* **1985**, *31*, 1770.

[114] Matthew, W.; Foulkes, C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.

[115] Koskinen, P.; Makinen, V. *Comp. Material Science* **2009**, *47*, 237.

[116] Witek, H.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2004**, *121*, 5163.

[117] Witek, H.; Morokuma, K. *J. Comp. Chem.* **2004**, *25*, 1858.

[118] Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. A* **2006**, *110*, 13551.

[119] Malolepsza, E.; Witek, H.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 237.

[120] Gaus, M.; Chou, C.; Witek, H.; Elstner, M. *J. Phys. Chem. A* **2009**, *113*, 11866.

[121] Bodrog, Z.; Aradi, B.; Frauenheim, T. *J. Chem. Theory Comput.* **2011**, *7*, 2654.

[122] Kazachkin, D.; Nishimura, Y.; Witek, H.; Irle, S.; Borguet, E. *J. Am. Chem. Soc.* **2011**, *133*, 8191.

[123] Li, W.; Irle, S.; Witek, H. *ACS Nano* **2010**, *4*, 4475.

[124] Malolepsza, E.; Lee, Y.; Witek, H.; Irle, S.; Lin, C.; Hsieh, H. *Int. J. Quant. Chem.* **2009**, *109*, 1999.

[125] Kaminski, S.; Gaus, M.; Phatak, P.; von Stetten, D.; Elstner, M.; Mroginski, M. *J. Chem. Theory Comput.* **2010**, *6*, 1240.

[126] Simon, A.; Rapacioli, M.; Mascetti, J.; Spiegelman, F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6771.

[127] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.

[128] Rapacioli, M.; Spiegelman, F.; Talbi, D.; Minerva, T.; Goursot, A.; Heine, T.; Seifert, G. *J. Chem. Phys.* **2009**, *130*, 244304.

[129] Li, J.; Zhu, T.; Cramer, C.; Truhlar, D. *J. Phys. Chem. A* **1998**, *102*, 1820.

[130] Kalinowski, J.; Lesnyg, B.; Thompson, J.; Cramer, C.; Truhlar, D. *J. Phys. Chem. A* **2004**, *108*, 2545.

[131] Trani, F.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 713.

[132] Barone, V.; Carnimeo, I.; Scalmani, G. *J. Chem. Theory Comput.* **2013**, *9*, 2052.

[133] Gaussian 09, Gaussian Inc., Wallingford, CT, 2009. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.

[134] Trani, F.; Scalmani, G.; Zheng, G.; Carnimeo, I.; Frisch, M.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 3304.

[135] Szabo, A., Ostlund, N. O., Eds. *MODERN QUANTUM CHEMISTRY. Introduction to Advanced Electronic Structure Theory;* Dover Publications, INC.: Mineola, New York, 1996.

[136] Watson, J. K. G. *Mol. Phys.* **1968**, *100*, 47.

[137] Watson, J. K. G. *Mol. Phys.* **1970**, *19*, 465.

[138] Miller, W. H.; Hernandez, R.; Handy, N. C.; Jayatilaka, D.; Willets, A. *Chem. Phys. Lett.* **1990**, *172*, 62.

[139] Vleck, J. H. V. *Phys. Rev.* **1929**, *33*, 467.

[140] Bloino, J.; Barone, V. *J. Chem. Phys.* **2012**, *136*, 124108.

[141] Clabo, D. A.; Allen, W. D.; Reminton, R. B.; Yamaguchi, Y.; Schaefer III, H. F. *Chem. Phys.* **1988**, *123*, 187.

[142] Schneider, W.; Thiel, W. *Chem. Phys. Lett.* **1989**, *157*, 367.

[143] Willets, A.; Handy, N. C.; Green, W. H.; Jayatilaka, D.

[144] Puzzarini, C.; Biczysko, M.; Barone, V. *J. Chem. Theory Comput.* **2010**, *6*, 828.

[145] Barone, V.; Bloino, J.; Guido, C. A.; Lipparini, F. *Chem. Phys. Lett.* **2010**, *496*, 157.

[146] Darling, B. T.; Dennison, D. M. *Phys. Rev.* **1940**, *57*, 128.

[147] Fermi, E. *Z. Physisik* **1931**, *71*, 251.

[148] Vibrationally-resolved electronic spectra in Gaussian 09. Barone, V.; Bloino, J.; Biczysko, M. **2009**.

[149] Raghavachari, K.; Trucks, G.; Pople, J.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

[150] Double and triple-ζ basis sets of sns and n07 families, are available for download. **2012**.

[151] Barone, V.; Cimino, P.; Stendardo, E. *J. Chem. Theory Comput.* **2008**, *4*, 751.

[152] Barone, V.; Cimino, P. *Chem. Phys. Lett.* **2008**, *454*, 139.

[153] Barone, V.; Cimino, P. *J. Chem. Theory Comput.* **2009**, *5*, 192.

[154] Bergner, A.; Dolg, M.; Kuchle, W.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *80*, 1431.

[155] Igelmann, G.; Stoll.; Preuss, H. *Mol. Phys.* **1988**, *65*, 1321.

[156] T.H. Dunning, J. *J. Chem. Phys.* **1989**, *90*, 1007.

[157] Kendall, A.; T.H. Dunning, J.; Harrison, R. *J. Chem. Phys.* **1992**, *96*, 6796.

[158] Woon, D.; T.H. Dunning, J. *J. Chem. Phys.* **1995**, *103*, 4572.

[159] Peterson, K.; Shepler, B.; Figgen, D.; Stoll, H. *J. Phys. Chem. A* **2006**, *110*, 13877.

[160] Peterson, K.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113.

[161] Balabin, R. *J. Phys. Chem. Lett.* **2010**, *1*, 20.

[162] Iijima, K.; Tanaka, K.; Onuma, S. *J. Mol. Struct.* **1991**, *246*, 257.

[163] Huisken, F.; Werhahn, O.; Ivanov, A.; Krasnokutski, S. *J. Chem. Phys.* **1999**, *111*, 7.

[164] Gomez-Zavaglia, A.; Fausto, R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3154.

[165] Matè, B.; Rodriguez-Lazcano, Y.; Galvèz, O.; Tanarro, I.; Escribano, R. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12268.

[166] Filler, M. A.; Bent, S. F. *Progress in surface science* **2003**, *73*, 1.

[167] Suenram, R. D.; Lovas, F. J. *J. Am. Chem. Soc.* **1980**, *102*, 7180.

[168] Barone, V.; Adamo, C.; Lelj, F. *J. Chem. Phys.* **1995**, *102*, 364.

[169] Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1991**, *113*, 7917.

[170] Csaszar, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9569.

[171] Kasalova, V.; Allen, W. D.; Schaefer III, H. F.; Czinki, E.; Csaszar, A. G. *J. Comp. Chem.* **2007**, *28*, 1373.

[172] Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1995**, *117*, 2019.

[173] Balabin, R. *Chem. Phys. Lett.* **2009**, *479*, 195.

[174] Meng, K.; Wang, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2001.

[175] Lopez, A.; Heller, T.; Bitzer, T.; Richardson, N. *Chem. Phys.* **2002**, *277*, 1.

[176] Qu, Y.; Wang, Y.; Li, J.; Han, K. L. **2004**, *569*, 12.

[177] Luo, X.; Qian, G.; Sagui, C.; Roland, C. *J. Phys. Chem. C* **2008**, *112*, 2640.

[178] Nguyen, D.; Scheiner, A.; Andzelm, J.; Sirois, S.; Salahub, D.; Hagler, A. *J. Comput. Chem.* **1997**, *18*, 1609.

[179] Hu, C. H.; Shen, M.; Schaefer, H. F. *J. Am. Chem. Soc.* **1993**, *115*, 2923.

[180] Bludsky, O.; Chocholousova, J.; Vacek, J.; Huisken, F.; Hobza, P. *J. Chem. Phys.* **2000**, *113*, 11.

[181] Shmilovits-Ofir, M.; Miller, Y.; Gerber, R. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8715.

[182] Ramstad, A.; Brocks, G.; Kelly, P. J. *Phys. Rev. B* **1995**, *51*, 14504.

[183] Besley, N. A.; Blundy, A. *J. Phys. Chem. B* **2006**, *110*, 1701.

[184] Northrup, J. *Phys. Rev. B* **1993**, *47*, 10032.

[185] Chadi, D. *Phys. Rev. Lett.* **1979**, *43*, 43.

[186] Borriello, I.; Cantele, G.; Ninno, D.; Iandonisi, G.; Cossi, M.; Barone, V. *Phys. Rev. B* **2007**, *76*, 035430.

[187] Festa, G.; Cossi, M.; Barone, V.; Cantele, G.; Ninno, D.; Iandonisi, G. *J. Chem. Phys.* **2005**, *122*, 184714.

[188] Radi, A.; Ebrahimi, M.; Leung, K. **2010**, *604*, 1073.

[189] Yoshinobu, J. *Progress in surface science* **2004**, *77*, 37.

[190] Waltenburg, H. N.; J. T. Yates, J. *Chem. Rev.* **1995**, *95*, 1589.

[191] Jung, Y. S.; Shao, Y. H.; Gordon, M. S.; Doren, D. J.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 10917.

[192] Jung, Y. S.; Akinaga, Y.; Jordan, K. D.; Gordon, M. S. *Theor. Chem. Acc.* **2003**, *109*, 268.

[193] Olson, R. M.; Gordon, M. S. *J. Chem. Phys.* **2006**, *124*, 081105.

[194] Shemesh, D.; Baer, R.; Seideman, T.; Gerber, R. *J. Chem. Phys.* **2005**, *122*, 184704.

[195] Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363.

[196] Hratchian, H. P.; Schlegel, H. B. *J. Chem. Theory Comput.* **2005**, *1*, 61.

[197] Widdra, W.; Huang, C.; Briggs, G. A. D.; Weinberg, W. H. *J. Electron Spectrosc. Relat. Phenom.* **1993**, *64*, 129.

[198] Niehaus, T. A.; Suhai, S.; Sala, F. D.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. *Phys. Rev. B* **2001**, *63*, 085108.

[199] Porezag, D.; Frauenheim, T.; Khler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947.

[200] Seifert, G.; Eschrig, H.; Bieger, W. *Z. Phys. Chem. (Leipzig)* **1986**, *267*, 529.

[201] Goringe, C. M.; Bowler, D. R.; Hernandez, E. *Rep. Prog. Phys.* **1997**, *60*, 1447.

[202] Menon, M.; Subbaswamy, K. R. *Phys. Rev. B* **1997**, *55*, 9231.

[203] Jancou, J. M.; Scholz, R.; Beltram, F.; Bassani, F. *Phys. Rev. B* **1998**, *57*, 6493.

[204] Ordejon, P.; Lebedenko, D.; Menon, M. *Phys. Rev. B* **1994**, *50*, 5645.

[205] Niehaus, T. A.; Della Sala, F. *Physica Status Solidi b-basic solid state physics* **2012**, *249*, 237.

[206] Kohler, C.; Seifert, G.; Gerstmann, U.; Elstner, M.; Overhof, H.; Frauenheim, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5109.

[207] Kohler, C.; Seifert, G.; Frauenheim, T. *Chem. Phys.* **2005**, *309*, 23.

[208] Zheng, G.; Witek, H.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Kohler, C.; Frauenheim, T. *J. Chem. Theory Comput.* **2007**, *3*, 1349.

[209] Janak, J. F. *Phys. Rev. B* **1978**, *18*, 7165.

[210] Moreira, N. H.; Dolgonos, G.; Aradi, B.; da Rosa, A. L.; Frauenheim, T. *J. Chem. Theory Comput.* **2009**, *5*, 605.

[211] Dolgonos, G.; Aradi, B.; Moreira, N. H.; Frauenheim, T. *J. Chem. Theory Comput.* **2010**, *6*, 266.

[212] Frisch, M.; Head-Gordon, M.; Pople, J. *Chem. Phys.* **1990**, *141*, 189.

[213] Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43.

[214] Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.

[215] Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

[216] Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1994**, *100*, 7495.

[217] Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1994**, *101*, 3888.

[218] Jamorski, C.; Casida, M.; Salahub, D. *J. Chem. Phys.* **1996**, *104*, 5134.

[219] Casida, M., Ed. *Recent advances in density functional methods (Part I). Chapter 5: Time-Dependent Density Functional Response Theory for Molecules.;* World Scientific Publishing Co. Pte. Ltd.: P O Box 128, Farrer Road, Singapore 912805, 1997.

[220] Cossi, M.; Barone, V. *J. Chem. Phys.* **2001**, *115*, 4708.

[221] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[222] Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

[223] Lu, Z.; Liu, H.; Elstner, M.; Yang, W. *Reviews of modern quantum chemistry: a celebration of the contribution of Robert G. Parr* **2002**, page 1606.

[224] Hou, G.; Zhu, X.; Cui, Q. *J. Chem. Theory Comput.* **2010**, *6*, 2303.

[225] Xie, L.; Liu, H. *J. Comp. Chem.* **2002**, *23*, 1404.

[226] Rapacioli, M.; Simon, A.; Dontot, L.; Spiegelman, F. *Phys. Stat. Sol. (b)* **2012**, *249*, 245.

[227] Giese, T.; York, D. *Theor. Chem. Acc.* **2012**, *131*, 1145.

[228] Giese, T.; York, D. *J. Chem. Phys.* **2011**, *134*, 194103.

[229] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2010**, *7*, 931.

[230] Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338.

[231] Kaminski, S.; Gaus, M.; Elstner, M. *J. Phys. Chem. A* **2012**, *116*, 11927.

[232] Kubar, T.; Elstner, M. *J. Phys. Chem. B* **2010**, *114*, 11221.

[233] Hourahine, B.; Sanna, S.; Aradi, B.; Niehaus, T.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5671.

[234] Lundberg, M.; Nishimoto, Y.; Irle, S. *Int. J. Quant. Chem.* **2011**, *112*, 1701.

[235] Begue, D.; Carbonniere, P.; Pouchan, C. *J. Phys. Chem. A* **2005**, *109*, 4611.

[236] Wiberg, K. B.; Thiel, Y.; Goodman, L.; Leszczynski, J. *J. Phys. Chem.* **1995**, *99*, 13850.

[237] Olbert-Majkut, A.; Ahokas, J.; Lundell, J.; Pettersson, M. *Journal of Raman Spectroscopy* **2011**, *42*, 1670.

[238] Tuck, P. O.; Mawhinney, R. C.; Rappon, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4471.

[239] Guido, C.; Mennucci, B.; Jacqueminc, D.; Adamo, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8016.

[240] Stancik, A. L.; Brauns, E. B. *Vibrational Spectroscopy* **2008**, *47*, 66.

List of Articles

[51] "Reliable structural, thermodynamics and spectroscopic properties of organic molecules adsorbed on silicon surfaces from computational modelling: the case of glycine@Si(100)"

I. Carnimeo, M. Biczysko, J. Bloino and V. Barone

Phys. Chem. Chem. Phys., **13**, 16713 (2011)

[134] "Time-Dependent Density Functional Tight Binding: New Formulation and Benchmark of Excited States"

F. Trani, G. Scalmani, G. Zheng, I. Carnimeo, M. J. Frisch and V. Barone

J. Chem. Theory Comput., **7**, 3304 (2011)

[50] "Toward anharmonic computations of vibrational spectra for large molecular systems"

V. Barone, M. Biczysko, J. Bloino, M. Borkowska-Panek, I. Carnimeo and P. Panek

Int. J. Quant. Chem., **112**, 2185 (2012)

[71] "Simulation of theoretical IR spectra for complex molecular systems from vibrational perturbative approaches: glycine as a test case"

M. Biczysko, J. Bloino, I. Carnimeo, P. Panek and V. Barone

J. Mol. Struct., **1009**, 74 (2012)

[132] "Computational Spectroscopy of Large Systems in Solution: The DFTB/PCM and TD-DFTB/PCM Approach"

V. Barone, I. Carnimeo and G. Scalmani

J. Chem. Theory Comput., **9**, 2052 (2013)

[70] "Anharmonic theoretical simulations of infrared spectra of halogenated organic compounds"

I. Carnimeo, C. Puzzarini, N. Tasinato, P. Stoppa, A. Pietropolli Charmet, M. Biczysko, C. Cappelli and V. Barone

J. Chem. Phys., **139**, 074310 (2013)

[73] "An integrated experimental and quantum-chemical investigation on the vibrational spectra of chlorofuoromethane"

A. Pietropolli Charmet, P. Stoppa, N. Tasinato, S. Giorgianni, V. Barone, M. Biczysko, J. Bloino, C. Cappelli, I. Carnimeo and C. Puzzarini

J. Chem. Phys., **139**, 164302 (2013)

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to prof. Barone, for the great scientific support he gave me in these years, always inducing me to improve my work and providing new ideas in the most difficult moments.

I wish to acknowledge also the other members of the group, in particular Malgorzata Biczysko, Julien Bloino and Chiara Cappelli, who guided me in writing the first articles and supported me during the development of my Ph.D. project. Thanks to their attention and availability I was able to learn a lot about spectroscopy.

I also would like to acknowledge the other people I worked with, writing many of the articles on which this thesis is based. To Giovanni Scalmani a sincere gratitude for his assistance in the issues connected with the theory and implementation of the DFTB method, both during my stay at the Gaussian Inc. and the other stages of the work. I also thank all the people met at the Gaussian Inc. for the kind welcome I received. I wish to acknowledge Cristina Puzzarini and the group of Molecular Spectroscopy of Ca' Foscari (Venezia) for the stimulating discussions about the computation of accurate spectroscopic properties of organic molecules, had in Bologna and in Pisa and during the writing of the related articles.

Furthermore, I would like to thank the people with whom I shared many moments of this Ph.D.: Filippo, for his friendship, his support during the early years and the very useful discussions about the PCM, Michele, Livia, Matteo, Mireia, Correntin, Franco and all the other students and post-docs of the group for their warm friendship and the pleasant time spent together.

Finally, a deep and sincere thank to my family for the precious and patient support they daily provided during this experience.